SVENSKA GEOFYSISKA FÖRENINGEN

VOLUME 4, NUMBER 3 **Tellus** AUGUST 1952

A QUARTERLY JOURNAL OF GEOPHYSICS

SYMPOSIUM ON NUMERICAL FORECASTING

Simplified Dynamic Models of the Atmosphere, designed for the Purpose of Numerical Weather Prediction

By ARNT ELIASSEN
Det Norske Meteorologiske Institutt, Oslo

(Manuscript received 5 August 1952)

Abstract

The properties of several atmospheric models, which have been proposed for the purpose of numerical computation of prognostic maps, are discussed, and a new version of the "$2^1/_2$-dimensional model" is given. It is pointed out that the "barotropic" model is unable to explain adequately the growth of wave perturbations, and that the advective model gives too strong instability for short waves, whereas the two-layer model and the continuous $2^1/_2$-dimensional model seem to describe almost correctly the behaviour of waves in a baroclinic current.

1. Introduction

Looking back at the results of the research on numerical weather forecasting carried out after the last war, we may say that a first stage of the research program is now finished, since it is definitely established that the two-dimensional, non-divergent model is able to explain a considerable part of the changes that we observe on the 500 mb maps. We have thus arrived at a serviceable first approximation, which enables us to carry on the research work with more confidence than before. The next stage, the advance from this bridgehead towards more accurate methods for numerical weather forecasting, has already begun. The goal is to find a system of prognostic equations which is realistic enough to give forecasts of practical value, and at the same time simple enough to permit computations of such forecasts with a reasonable amount of labour. Since this is a compromise, there may not be just one solution, suitable for all purposes. We may perhaps need several methods, of various degrees of accuracy and mathematical complexity; which one to use will depend on the nature of the forecast in question, as well as on the capacity of the available calculating equipment.

The derivation of suitable prognostic equations, which fulfil the conditions mentioned above, is one of the two main problems of numerical weather forecasting. We may hope to arrive at such equations by testing a variety of intelligent proposals. A discussion of some of the models that have been proposed, together with the author's version of the "2 ½-dimensional model", is given below. The second main problem of numerical forecasting, to overcome the mathematical difficulties

involved in the process of numerical integration, will not be discussed here.

2. Quasi-geostrophic equations in three dimensions

So far, most of the prognostic equations that have been considered have been based on the quasi-static and the quasi-geostrophic approximations. The quasi-static approximation is valid for motions whose horizontal extent is large in comparison with their vertical extent, and is generally believed to be legitimate for all large-scale motions in the atmosphere. More questionable is the legitimacy of the quasi-geostrophic approximation, which has been discussed by CHARNEY (1948). If we study solutions of linearized equations, we find that this approximation, together with the quasi-static approximation, eliminates all sound waves and ordinary gravitational-inertial waves whose perturbation velocities are completely non-geostrophic. On the other hand, the same equations describe almost correctly the behaviour of waves with nearly geostrophic perturbation velocities, such as the gravitational-inertial waves in a baroclinic current discussed by CHARNEY (1947) and EADY (1949), as well as Rossby-waves. Since most large-scale disturbances connected with weather seem to belong to the group with almost geostrophic perturbation velocities, we have reason to hope that the quasi-geostrophic approximation will serve our purpose. On the other hand, we know from synoptic experience that large non-geostrophic winds sometimes exist in the upper troposphere; in particular, we must expect severe difficulties when applying the geostrophic approximation to sharp upper ridges, whose importance for the dynamics of the flow has been stressed by J. BJERKNES (1951). We must therefore expect the quasi-geostrophic approximation to cause considerable errors in some cases.

The quasi-static approximation enables us to use pressure (p) instead of height (z) as a vertical coordinate, which is convenient for several reasons. This means that partial differentiation with respect to horizontal coordinates (x, y) and time (t) will be performed at constant pressure, such that $\partial/\partial x$ and $\partial/\partial y$ denote differentiation along an isobaric surface, and $\partial/\partial t$ means rate of change as measured by an observer which follows an isobaric surface in its vertical motion.

It is convenient to use vector notation $\mathbf{v}$ for the horizontal wind velocity (with components u and v), and to introduce a horizontal del-operator ∇ (with components $\partial/\partial x$ and $\partial/\partial y$) to express differentiation with respect to horizontal coordinates (at constant p). Let further $\mathbf{k}$ denote a vertical unit vector, pointing upward; ω individual rate of change of pressure, α specific volume, Θ potential temperature, f Coriolis parameter and g acceleration of gravity.

The hydrostatic equation may then be written

$$\frac{\partial z}{\partial p} = -\frac{\alpha}{g}. \tag{1}$$

The equation of continuity assumes the simple form

$$\nabla \cdot \mathbf{v} + \frac{\partial \omega}{\partial p} = 0. \tag{2}$$

Neglecting frictional forces, we derive from the horizontal equation of motion the vorticity equation

$$\frac{\partial \zeta}{\partial t} + \mathbf{v} \cdot \nabla (f + \zeta) + \omega \frac{\partial \zeta}{\partial p} + \\ + \mathbf{k} \cdot \nabla \omega \times \frac{\partial \mathbf{v}}{\partial p} + (f + \zeta) \nabla \cdot \mathbf{v} = 0, \tag{3}$$

where

$$\zeta = \mathbf{k} \cdot \nabla \times \mathbf{v} = \frac{\partial v}{\partial x} - \frac{\partial u}{\partial y} \tag{4}$$

is the "isobaric" vorticity, which differs from the vertical vorticity component in the differentiations being taken along isobaric instead of level surfaces.

Elimination of $\nabla \cdot \mathbf{v}$ between (2) and (3) gives the vorticity equation in the form

$$\frac{\partial \zeta}{\partial t} + \mathbf{v} \cdot \nabla (f + \zeta) + \omega \frac{\partial \zeta}{\partial p} + \\ + \mathbf{k} \cdot \nabla \omega \times \frac{\partial \mathbf{v}}{\partial p} - (f + \zeta) \frac{\partial \omega}{\partial p} = 0. \tag{5}$$

Since the amount of heating and cooling is not known, we are at the present time

unable to take this effect into account, and will have to assume that the potential temperature is conserved. The equation expressing this assumption is

$$(6) \qquad \frac{1}{\alpha}\frac{\partial \alpha}{\partial t} + \frac{1}{\alpha}\mathbf{v}\cdot\nabla\alpha + \frac{\omega}{\Theta}\frac{\partial \Theta}{\partial p} = 0,$$

since differentiation at constant pressure gives $\partial \ln \Theta/\partial t = \partial \ln \alpha/\partial t$ and $\nabla \ln \Theta = \nabla \ln \alpha$. Making use of the hydrostatic equation (1) in the first term, this equation may also be written

$$(7) \qquad \frac{\partial}{\partial p}\frac{\partial z}{\partial t} - \frac{1}{g}\mathbf{v}\cdot\nabla\alpha - \frac{\alpha\omega}{g\Theta}\frac{\partial \Theta}{\partial p} = 0.$$

The quasi-geostrophic method consists in substituting the geostrophic wind for the true wind in the terms which are not critically affected by this substitution. Thus the geostrophic wind approximation cannot be used in the divergence term, or in the expression for the work of the horizontal pressure force (if the energy equations are considered). The simplest and most logical way to introduce the geostrophic approximation is given by CHARNEY (1948). His method consists in substituting the geostrophic wind for the true wind everywhere in the thermodynamic energy equation and in the vorticity equation, after having eliminated the horizontal divergence by means of the equation of continuity.

From the geostrophic wind formula

$$(8) \qquad \mathbf{v}_g = \frac{g}{f}\mathbf{k}\times\nabla z$$

we find the following approximate formula for the geostrophic vorticity

$$(9) \qquad \zeta_g = \frac{g}{f}\nabla^2 z = \frac{g}{f}\left(\frac{\partial^2 z}{\partial x^2} + \frac{\partial^2 z}{\partial y^2}\right).$$

From these equations and the hydrostatic equation (1) we find

$$(10) \qquad \frac{\partial \mathbf{v}_g}{\partial p} = -\frac{1}{f}\mathbf{k}\times\nabla\alpha,$$

$$(11) \qquad \frac{\partial \zeta_g}{\partial p} = -\frac{1}{f}\nabla^2\alpha.$$

Using Charney's method, we now introduce the geostrophic wind approximation into equations (5) and (7), and obtain

$$(12) \qquad \nabla^2\frac{\partial z}{\partial t} - \frac{f(f+\zeta_g)}{g}\frac{\partial \omega}{\partial p} - \frac{1}{g}\nabla\cdot(\omega\nabla\alpha) = -\mathbf{k}\cdot\nabla z\times\nabla(f+\zeta_g),$$

$$(13) \qquad \frac{\partial}{\partial p}\frac{\partial z}{\partial t} - \frac{\alpha}{g\Theta}\frac{\partial \Theta}{\partial p}\omega = \frac{1}{f}\mathbf{k}\cdot\nabla z\times\nabla\alpha.$$

This may be considered as a system of linear differential equations for the two unknown functions $\partial z/\partial t$ and ω, whose coefficients can be determined if the field $z\,(x, y, p)$ of the contour height is known at the instant considered. When the absolute vorticity and the static stability are both positive, which corresponds to normal conditions in the atmosphere, the system is of the elliptic type. Together with the boundary conditions at the top and the bottom of the atmosphere

$$(14) \qquad \begin{cases} \omega = 0 & \text{at } p = 0, \\ \dfrac{\partial z}{\partial t} - \dfrac{1}{g}\alpha\omega = 0 & \text{at the ground, assumed to be level } (p = p_0 \approx 1000 \text{ mb}) \end{cases}$$

and suitable conditions for the lateral boundaries, the system (12, 13) determines $\partial z/\partial t$ and ω throughout the region considered. The solution may be worked out numerically, and may be used to extrapolate the field $z(x, y, p)$ over a small interval of time. Iterating this procedure, we obtain an integration in small steps of time. This method, based on essentially the same system of equations, has been discussed previously by the author (ELIASSEN, 1949).

In order to solve the system (12, 13) numerically, we may eliminate one of the unknown functions, whence we obtain a Poisson-type differential equation in three dimensions for the other, which may be solved e. g. by relaxation. Such an equation for the height (or pressure) tendency has been discussed by CHARNEY (1949).

It is interesting to note that if the third term on the left of (12) is neglected (which is often done), then the system (12), (13) is equivalent to the equilibrium equations for a loaded elastic system consisting of a number

of horizontal elastic membranes exposed to weights of variable magnitude, and connected with vertical springs. Interpreting $\partial z/\partial t$ as the vertical displacement of the membranes and ω as the tension in the vertical springs, we see that the equation expressing equilibrium between the vertical forces has the form (12), whereas (13) expresses a linear stress-strain relationship for the vertical springs. The boundary conditions (14) also have obvious interpretations in this elastic model. Since we have an immediate feeling of how this model will behave under the influence of a given distribution of weights, and for given values of the constant term in Hooke's law (right-hand side of (13)), it may serve to give us an idea of what the solutions of our meteorological problem will look like.

Computations of forecasts based on the three-dimensional, quasi-geostrophic equations involve a large amount of computational work. To the author's knowledge, no such forecasts have yet been computed. So far, the tests have been based on simplified equations, which reduce the problem to one in two dimensions in space. This is attained by making certain assumptions, based on empirical knowledge, of the variation with height of the field of motion.

3. The barotropic model

The simplest model of this kind is the so-called barotropic model. The suggestion that the large-scale motion of the atmosphere may be considered as two-dimensional and non-divergent was first put forward by ROSSBY (1939). CHARNEY (1949) has given a more precise formulation of the physical basis for this hypothesis. He uses the vorticity equation (5) in the simplified form

$$(15) \qquad \frac{\partial \zeta}{\partial t} + \mathbf{v} \cdot \nabla (f + \zeta) - f \frac{\partial \omega}{\partial p} = 0$$

with the simplified boundary conditions

$$(16) \qquad \omega = 0 \text{ at } p = 0 \text{ and at } p = p_0.$$

Furthermore, he makes use of the observed fact that as far as the large-scale motion is concerned, the magnitude of the wind velocity tends to change with height in a typical fashion, whereas the streamline pattern stays about the same from level to level. This assumption makes it possible to carry out an integration of (15) with respect to p from 0 to p_0, thus eliminating ω. The resulting equation expresses that at a certain intermediate level, which is found to be in the vicinity of 500 millibars, vorticity changes are due to advection only:

$$(17) \qquad \frac{\partial \zeta}{\partial t} = -\mathbf{v} \cdot \nabla (f + \zeta) \text{ at } p \approx 500 \text{ millibars.}$$

The motion at this level may therefore be considered as horizontal and non-divergent. When the geostrophic assumption is introduced, a two-dimensional Poisson equation for the height tendency is obtained, which may be used as a prognostic equation[1].

Several numerical tests have been carried out by comparing the changes computed from this equation with the observed ones (CHARNEY, ELIASSEN, 1949; CHARNEY, FJØRTOFT, VON NEUMANN, 1950; BOLIN, CHARNEY, 1951; STAFF MEMBERS, UNIVERSITY OF STOCKHOLM, 1952). The results are very promising indeed; the discrepancies, although important, are nothing more than we would have to expect from such a crude model.

The properties of horizontal, two-dimensional flow, which is governed by equation (17), has been subject to many theoretical studies. ROSSBY (1939) investigated horizontal, non-divergent wave perturbations on a uniform zonal current, and found that these waves move with the speed

$$(18) \qquad c = U - \frac{\beta}{k^2}$$

where U is the velocity of the current, $\beta = df/dy$, and k the wave number. These waves are stable for all wave lengths. CRAIG (1945) and HØILAND (1950) have found solutions representing two-dimensional non-divergent waves with finite amplitude on a spherical earth. These theoretical motions have a great resemblance with the observed wave patterns on the 500 mb maps.

The most apparent defect of the barotropic model seems to be connected with the growth

[1] Concerning the details, the reader is referred to CHARNEY's paper (1949) and also to CHARNEY and ELIASSEN (1949), CHARNEY, FJØRTOFT, VON NEUMANN (1950).

of wave perturbations. Unstable waves do exist in the barotropic model, but the only source of energy for the growing perturbations is the kinetic energy of the basic flow. Investigations made by CHARNEY (1947), EADY (1949) and FJØRTOFT (1950), of wave perturbations in a baroclinic, zonal current, based on the three-dimensional equations, have shown the existence of unstable waves that grow at the expense of the potential energy of the system; and there is rather strong evidence that most waves in the westerlies develop as a result of instability of this kind. Such unstable waves do not exist in the barotropic model, since transformations of potential into kinetic energy are excluded a priori in this model. We must conclude that the barotropic model is not capable of an adequate explanation of the growth of wave perturbations, and we must therefore look for more realistic models.

4. The advective model

It is possible to take into account important effects related to transformation of potential into kinetic energy, i. e. to the baroclinicity, without using the complicated three-dimensional equations. So called 2 ½-dimensional models have been designed, which in a crude manner take into account the three-dimensional structure of the atmosphere and the vertical motion, although they are still two-dimensional, mathematically speaking. In these models, two horizontal fields (e. g. one contour field and one temperature field) are being used to define the state and motion of the atmosphere. A better name would therefore perhaps be "two-parametric models".

The simplest model of this type is *the advective model*. It is based on the assumption that temperature changes are caused by horizontal advection only. Moreover, one assumes that the thermal wind has a similar variation in magnitude with height in all places, whereas its direction is independent of height. In other words, one assumes that the wind field can be represented by the formula

$$(19)\quad \mathbf{v}(x, y, p, t) = \overline{\mathbf{v}}(x, y, t) + A(p)\,\mathbf{v}_T(x, y, t)$$

where the bar denotes the mean value with respect to pressure, $A(p)$ represents the variation of thermal wind with height, and $\mathbf{v}_T$ is the thermal wind measured between two selected isobaric levels. Introducing the thickness h between these isobaric levels, and the mean contour height $\bar{z}$, the geostrophic assumption gives

$$(20)\quad \overline{\mathbf{v}} = \frac{g}{f}\,\mathbf{k}\times\nabla\bar{z}, \quad \mathbf{v}_T = \frac{g}{f}\,\mathbf{k}\,x\nabla h,$$

and the corresponding vorticities

$$(21)\quad \bar{\zeta} = \frac{g}{f}\,\nabla^2\bar{z}, \quad \zeta_T = \frac{g}{f}\,\nabla^2 h.$$

When these expressions are introduced into the approximate vorticity equation (15), and the mean value with respect to pressure is taken, one obtains

$$(22)\quad \frac{\partial\bar{\zeta}}{\partial t} = -\,\overline{\mathbf{v}}\cdot\nabla(f+\bar{\zeta}) - \overline{A^2}\,\mathbf{v}_T\cdot\nabla\zeta_T,$$

where the last term gives the contribution of the thermal wind field to the mean vorticity change. The advective hypothesis gives finally

$$(23)\quad \frac{\partial h}{\partial t} = -\,\overline{\mathbf{v}}\cdot\nabla h.$$

The two last equations are seen to contain the two horizontal fields $\bar{z}$ and h only, and may be integrated numerically in steps of time to give the change of these fields.

In a form suitable for numerical integration, the advective model is described in a paper by CHARNEY, FJØRTOFT and VON NEUMANN (1950), (see also FJØRTOFT, 1951, and CHARNEY, 1951). The assumption that temperature changes are due to horizontal advection only has actually been used earlier by ROSSBY (1942) and by J. BJERKNES and HOLMBOE (1944) in the study of long waves in a baroclinic zonal current, although without the derivation of the corresponding formula for the wave velocity. If we linearize (22) and (23) for the case of small perturbations on a baroclinic zonal current without horizontal shear, we find for sine waves of wave number k and of infinite lateral extent the velocity of propagation

$$(24)\quad c = \overline{U} - \frac{\beta}{2k^2} \pm \sqrt{\left(\frac{\beta}{2k^2}\right)^2 - \overline{A^2}\,U_T^2}$$

where $\overline{U}$ is the mean wind speed and U_T the thermal wind of the current. This formula has been derived by FJØRTOFT (1950) and by SUTCLIFFE (1951). A similar formula is given by BERSON (1950).

Equation (24) gives stable or unstable waves, according as the wave length is greater or smaller than a certain critical wave length, the magnitude of which increases with the square root of the thermal wind. This is in agreement with results derived by CHARNEY (1947) from the three-dimensional equations. Thus the advective model seems to be more realistic than the barotropic model, which in this case gives stable waves only. For wave lengths considerably larger than the critical wave length, one of the values (24) of the wave speed is in approximate agreement with the Rossby formula (18), whereas the other is considerably larger. This second value, however, corresponds to waves with a nodal surface in the middle of the atmosphere, roughly at the level where the barotropic model applies. These waves are therefore eliminated a priori in the barotropic model. We may conclude that the barotropic model is in agreement with the advective model as far as waves considerably longer than the critical wave length are concerned. For waves comparable with, or shorter than the critical wave length, the results derived from the two models differ appreciably (compare SUTCLIFFE 1951).

The unstable waves derived from the advective model have the property that their growth-rate increases indefinitely as the wave length approaches zero. This "ultra-violet catastrophe" has the unagreeable consequence that we are unable to solve the linearized initial value problem by means of Fourier analysis; most likely, this problem has no meaning. We may find a solution by dropping the shortest wave components, which is automatically done if the continuous functions are replaced by their values in a grid of points. However, such a procedure will not eliminate the difficulty, because the solution will depend critically on the wavelength of the shortest waves kept in the computation, i. e. on the grid size. We must expect to find the same difficulty also if we apply the equations of the advective model in non-linear form.

The extreme instability for very short waves, obtained from the advective model, is not supported by the theory based on the three-dimensional equations. This instability is just an error, introduced by the advective assumption, which is not legitimate for short waves. As we shall see, the error is quite unnecessary, and the advective model is therefore not a particularly good model to use for the purpose of numerical forecasting.

We shall now proceed to discuss some two-parametric models where also the vertical transport of potential temperature is taken into account.

5. Phillips' two-layer model

PHILLIPS (1951) has designed a model consisting of two homogeneous and incompressible fluid layers separated by a quasi-horizontal interface and bounded by two rigid, horizontal planes. The density of the upper layer is somewhat smaller than that of the lower, so that the system is statically stable. The system is baroclinic if the isobaric surfaces intersect the interface; and since each layer is autobarotropic, the baroclinicity is concentrated in the interface. The interface is permitted to slope and to move up or down, corresponding to vertical motion and horizontal divergence in each layer.

From the hydrostatic relation and the dynamical boundary condition at the interface one derives a linear relationship between the contour heights $z(x, y, t)$ and $z'(x, y, t)$ of the two layers and the height $h(x, y, t)$ of the interface. The state of the system is thus defined by two two-dimensional fields, e. g. the fields of z and h. Application of the geostrophic potential vorticity equation to the two layers gives for these two fields differential equations of the form

$$(25) \qquad \nabla^2 \frac{\partial h}{\partial t} - a(x, y) \frac{\partial h}{\partial t} = q(x, y)$$

$$(26) \qquad \nabla^2 \frac{\partial z}{\partial t} = r(x, y)$$

where the coefficient functions a and q can be computed from knowledge of the fields h and z, and r can be computed from knowledge of h, z and $\partial h/\partial t$.[1] Together with

[1] For further details see PHILLIP's paper (1951).

suitable lateral boundary conditions, these equations may be used to determine $\partial h/\partial t$ and $\partial z/\partial t$ when z and h are known at the instant considered. The tendencies thus computed may be used to extrapolate z and h through a finite interval of time; and iteration of this procedure gives a numerical integration in steps of time. It is important to note that although the motion of the model is three-dimensional, the forecast problem is mathematically two-dimensional in space, since the vertical coordinate (p) does not occur as an independent variable.

Phillips has applied his model to study the behaviour of perturbations on a uniform zonal current with different wind speeds in the two layers. For long waves, his results are in agreement with Charney's results and with the advective model: waves longer than a critical wave length are stable, and shorter waves unstable. However, as the wave length decreases, the growth rate of the unstable waves does not increase indefinitely, as in the advective model, but instead increases to a maximum and then falls down to zero at another critical wave length. Still shorter waves are stable in Phillips' model. This part of the frequency curve is in agreement with the results obtained by EADY (1949) from the three-dimensional perturbation equations for small wave perturbations on a baroclinic current. The whole frequency curve for the two-layer model corresponds closely to the frequency curve for a continuous medium derived by FJØRTOFT (1950) from energy considerations. This is a strong indication that the two-layer model in many respects will behave like the three-dimensional atmosphere.

Phillips has also applied his model to an actual weather situation and computed tendencies and vertical velocities which look very promising.

A disadvantage of the two-layer model is its geometrical dissimilarity with the real atmosphere, which makes it difficult to define a proper correspondence between the states of the model and the states of the atmosphere. Phillips establishes such a correspondence by requiring the frequency curve, derived from the model for waves on a zonal current, to be quantitatively correct in its main properties. But even with the best possible choice of correspondence, one would expect the model to reproduce atmospheric motion with a certain amount of distortion. For the purpose of numerical forecasting, a more life-like model would therefore probably give better results.

6. Continuous two-parametric model with vertical stability

The continuous two-parametric model with static stability is a natural generalization of the barotropic and advective models, and it is therefore not surprising that several research workers seem to have arrived at this model more or less independently. Thus Eady has proposed a very simple and elegant version of this model which he presents in an article in this issue of Tellus (cf pp 157—167).

In the author's version of the 2 ½-dimensional model, described below, it has been attempted to attain the closest possible correspondence with atmospheric conditions, in order to minimize errors due to geometrical dissimilarity between the atmosphere and the model. The author's model is somewhat more complicated than Eady's, but does not differ much from Eady's model in its main properties.

In the two-parametric model, the temperature distribution along the vertical is represented by one parameter. In other words, the possible ascent curves of our model are restricted to a one-dimensional mannifold of curves. Moreover, it will be convenient to have the temperature (T) as a linear function of the parameter defining the particular ascent curve. We will therefore make the following assumption for the vertical temperature distribution:

$$T(p) = T_n(p) + \tau(p)\,(h - h_n). \tag{27}$$

Here h is the parameter which defines the temperature curve, and h_n a constant included for the sake of convenience. The function $T_n(p)$ is chosen as a "normal" ascent curve (fig. 1), and $\tau(p)$ is chosen positive in the troposphere and negative in the stratosphere, as shown in fig. 2. Formula (27) then gives, for different values of h, a family of ascent curves as shown in fig. 3. These curves may be thought of as a set of homologes for different air masses; they are in accordance with real conditions in so far as a warm troposphere

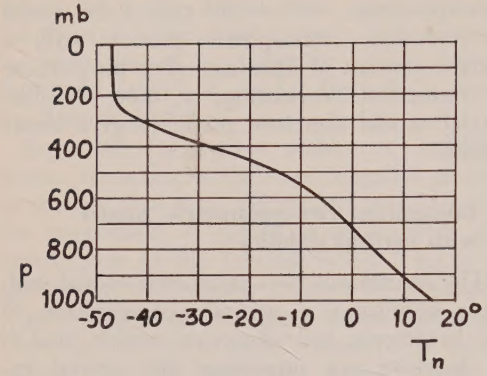


Fig. 1. "Normal" ascent curve T_n.

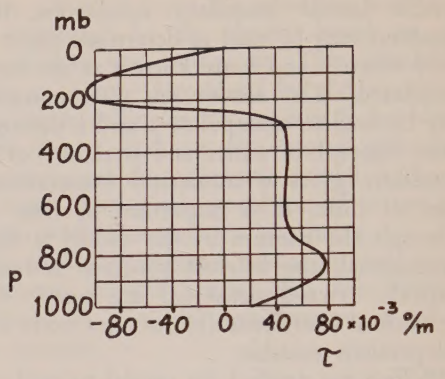


Fig. 2. Temperature deviation τ per unit thickness.

goes together with a cold stratosphere and vice versa. Since the result is unchanged if τ is reduced and at the same time h and h_n increased by the same factor, it follows that we may always choose h to represent the thickness of a prescribed isobaric layer, e. g. from 1,000 to 500 mb; h_n will then be the corresponding thickness for the "normal" temperature curve T_n. Assuming h to be a function of x, y and t, formula (27) will represent a three-dimensional temperature field.

The corresponding field of specific volume is

$$(28) \qquad \alpha(x, y, p, t) = R\frac{T_n(p)}{p} + R\frac{\tau(p)}{p}[h(x, y, t) - h_n],$$

(R, gas constant), and the contour field

$$(29) \qquad z(x, y, p, t) = \bar{z}(x, y, t) + H(p) + A(p)[h(x, y, t) - h_n].$$

Here the bar denotes as before the mean value with respect to pressure, and the functions $H(p)$ and $A(p)$ are defined by

$$(30) \qquad H' = -\frac{RT_n(p)}{gp}, \qquad \bar{H} = 0,$$

$$(31) \qquad A' = -\frac{R\tau(p)}{gp}, \qquad \bar{A} = 0.$$

The three-dimensional field $z(x, y, p, t)$ thus depends on the two horizontal fields of $\bar{z}$ and h.

From (29) we find the field of geostrophic wind

$$(31) \qquad \mathbf{v}(x, y, p, t) = \bar{\mathbf{v}}(x, y, t) + A(p)\,\mathbf{v}_T(x, y, t),$$

and the corresponding geostrophic vorticity

$$(32) \qquad \zeta(x, y, p, t) = \bar{\zeta}(x, y, t) + A(p)\,\zeta_T(x, y, t),$$

where $\bar{\mathbf{v}}$, $\mathbf{v}_T$, $\bar{\zeta}$, and ζ_T are defined by (20) and (21). Fig. 4 shows the function $A(p)$.

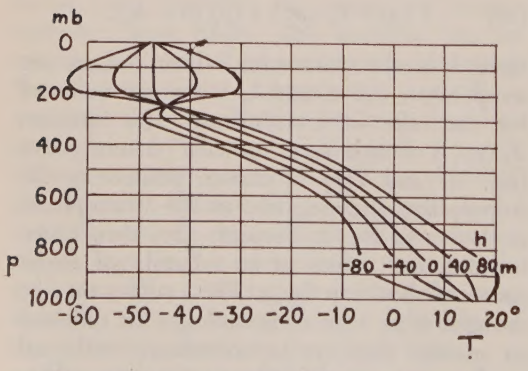


Fig. 3. Ascent curves for different values of h.

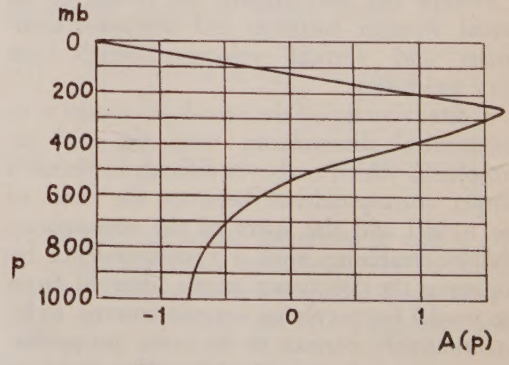


Fig. 4. The function $A(p)$.

We will also have to make a certain assumption about the distribution with height of vertical motion, and put

$$(33) \quad \omega\,(x, y, p, t) = B\,(p) \cdot w\,(x, y, t)$$

where $B\,(p)$ is chosen as a function of the type shown in fig. 5. The approximate boundary conditions (16) are automatically fulfilled when B is chosen so that $B\,(0) = B\,(p_0) = 0$.

We now insert the formulae (28), (31) and (33) into the thermodynamic equation (6). Approximating α, when undifferentiated, as well as the static stability factor, by their normal values, we obtain

$$(34) \quad \frac{\tau}{T_n}\left(\frac{\partial h}{\partial t} + \bar{\mathbf{v}} \cdot \nabla h\right) + \frac{B}{\Theta_n}\frac{d\Theta_n}{dp}\, w = 0,$$

where Θ_n is the potential temperature corresponding to the "normal" temperature T_n. As a rule, we cannot attain this equation to be exactly fulfilled at all levels; a reasonable requirement is, however, that the mean value of its left hand side shall vanish. This gives

$$(35) \quad w = -\gamma\left(\frac{\partial h}{\partial t} + \bar{\mathbf{v}} \cdot \nabla h\right),$$

where we have put

$$(36) \quad \gamma = \left(\overline{\frac{\tau}{T_n}}\right) \cdot \left[\overline{\left(\frac{B}{\Theta_n}\frac{d\Theta_n}{dp}\right)}\right]^{-1}.$$

The constant γ can be evaluated when the functions T_n, τ and B have been chosen.

The vorticity equation shall be used in the form

$$(37) \quad \frac{\partial \zeta}{\partial t} + \mathbf{v} \cdot \nabla\,(f + \zeta) + \omega\frac{\partial \zeta}{\partial p} - \\ - (f + \zeta)\frac{\partial \omega}{\partial p} = 0,$$

where only the term involving $\nabla\omega$ of the "exact" equation (5) has been dropped. It is even possible to take this term into account, but that would complicate the computations considerably. Introduction of (31), (32) and (33) into (37) gives

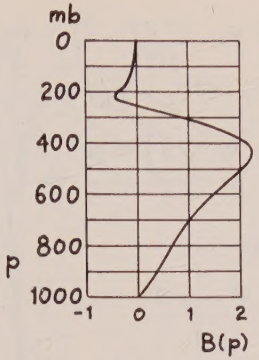


Fig. 5. The function $B\,(p)$.

$$(38) \quad \frac{\partial \bar{\zeta}}{\partial t} + A\frac{\partial \zeta_T}{\partial t} + \bar{\mathbf{v}} \cdot \nabla\,(f + \bar{\zeta}) + \\ + A\,[\bar{\mathbf{v}} \cdot \nabla \zeta_T + \mathbf{v}_T \cdot \nabla\,(f + \bar{\zeta})] + \\ + A^2\,\mathbf{v}_T \cdot \nabla \zeta_T - B'\,(f + \bar{\zeta})\,w + \\ + (A'B - AB')\,\zeta_T\,w = 0.$$

As in the case of the thermodynamic equation (34), we cannot hope to have the vorticity equation exactly satisfied at all levels. Our model is thus inconsistent, a fact which, however, will not disturb a theoretical meteorologist. We have the freedom to impose two conditions on the left-hand side of the vorticity equation, because we need two prognostic equations for the two horizontal fields $\bar{z}$ and h. What we have to do is to choose these conditions so as to minimize the error in the vorticity equation.

For fixed values of x, y and t, the left-hand side of (38) is a function of p, which will be denoted by $M(p)$. To this function we define a linear regression function $L(p)$, such that $\overline{(L - M)^2}$ = a minimum. We now require $L(p) \equiv 0$ for all values of p. This gives us the two wanted conditions

$$(39) \quad \bar{M} = 0 \text{ and } \overline{pM} - \bar{p} \cdot \bar{M} = 0.$$

When we here insert for M the left-hand side of (38), carry out the details of the calculation and eliminate w by means of (35), we obtain the two prognostic equations

$$(40) \quad \frac{\partial \bar{\zeta}}{\partial t} + \bar{\mathbf{v}} \cdot \nabla\,(f + \bar{\zeta}) + a\,\mathbf{v}_T \cdot \nabla \zeta_T - \\ - q\,\zeta_T\left(\frac{\partial h}{\partial t} + \bar{\mathbf{v}} \cdot \nabla h\right) = 0,$$

$$(41)\quad \frac{\partial \zeta_T}{\partial t} + \mathbf{v}_T \cdot \nabla (f + \bar{\zeta}) + \bar{\mathbf{v}} \cdot \nabla \zeta_T + \\ + b\, \mathbf{v}_T \cdot \nabla \zeta_T - r (f + \bar{\zeta} + A^\star \zeta_T) \\ \left(\frac{\partial h}{\partial t} + \bar{\mathbf{v}} \cdot \nabla h\right) = 0,$$

where

$$(42)\quad \begin{cases} a = \overline{A^2} \\ b = \dfrac{\overline{pA^2} - \bar{p}\,\overline{A^2}}{\overline{pA}} \\ q = 2\gamma\, \overline{A'B} \\ r = \dfrac{\gamma\, \bar{B}}{\overline{pA}} \\ A^\star = \dfrac{1}{\bar{B}} \left(\overline{AB} + 2\,\overline{pA'B} - 2\bar{p}\,\overline{A'B}\right). \end{cases}$$

Equations (40) and (41) represent, when the geostrophic formulae (20) and (21) are inserted, two linear differential equations for the tendencies $\partial \bar{z}/\partial t$ and $\partial h/\partial t$, of the same type as the prognostic equations (25), (26) obtained for the two-layer model. The system (40), (41) may be integrated numerically in steps of time. At each step, we will first have to determine $\partial h/\partial t$ from (41); having done that, $\partial \bar{z}/\partial t$ may be obtained from (40). Both equations may be solved by relaxation. If desired, the contour height at any particular level (e. g. 500 mb) may be used instead of the mean height $\bar{z}$, the relation between them being given by (29). This has the advantage that the two dependent variables h and z (500) can then be compared directly with the synoptic maps.

Each step will involve the evaluation of the fields of five advection terms, which have the form of Jacobian determinants. Their number may, however, be reduced to three. Let

$$(43)\quad J_1 = (\bar{\mathbf{v}} + A_1 \mathbf{v}_T) \cdot \nabla (f + \bar{\zeta} + A_1 \zeta_T), \\ J_2 = (\bar{\mathbf{v}} + A_2 \mathbf{v}_T) \cdot \nabla (f + \bar{\zeta} + A_2 \zeta_T)$$

denote the vorticity advection at two levels p_1 and p_2, which are chosen such that $A_1 = A(p_1)$ and $A_2 = A(p_2)$ are the roots of the equation

$$(44)\quad A^2 - bA - a = 0.$$

Then we have

$$(45)\quad \bar{\mathbf{v}} \cdot \nabla (f + \bar{\zeta}) + a \mathbf{v}_T \cdot \nabla \zeta_T = \frac{A_2 J_1 - A_1 J_2}{A_2 - A_1},$$

$$(46)\quad \mathbf{v}_T \cdot \nabla (f + \bar{\zeta}) + \bar{\mathbf{v}} \cdot \nabla \zeta_T + \\ + b\, \mathbf{v}_T \cdot \nabla \zeta_T = \frac{J_1 - J_2}{A_1 - A_2},$$

which we may introduce into (40) and (41).

The coefficients of the prognostic equations (40), (41) may be computed from (42) and (36), when the functions T_n, A and B have been chosen. If we adopt the values given in figs. 1, 4 and 5, which are based on estimates made by the author's colleague, Mr. Smebye, we find

$$(47)\quad \begin{aligned} a &= 0.65, & b &= 0.74, \\ q &= 0.89 \cdot 10^{-2}\, m^{-1}, & r &= 0.76 \cdot 10^{-2}\, m^{-1}, \\ A^\star &= 0.85. \end{aligned}$$

However, there is no unique way of defining the functions A and B, and the above values of the coefficients may therefore not be the best possible ones.

To the author's knowledge, all two-parametric models designed so far give the same prognostic equations (40), (41), only with different numerical values of the coefficients. This applies to the model proposed by Eady, as well as to Phillips' two-layer model, when suitably interpreted. For the purpose of numerical forecasting, it is therefore not necessary to pay too much attention to the details of the structure of the model; the best approach is perhaps to determine the coefficients of (40), (41) empirically so as to get the least error in the computed forecast.

7. Waves in a baroclinic zonal current.

We shall finally investigate on the basis of the continuous two-parametric model the speed of propagation and growth-rate of wave perturbations in a baroclinic zonal current. Let the wind velocity of the basic current at any level be given by $\bar{U} + A(p)\, U_T$, where $\bar{U}$ and U_T are constants. For simplicity, we shall let our coordinate system move with the mean wind, so that we may assume $\bar{U} = 0$. Assuming the perturbations $\bar{z}$

and h of the mean height and thickness to be functions of x and t only, we find from (40) and (41) the perturbation equations:

$$(48) \qquad \frac{\partial^3 \bar{z}}{\partial x^2 \partial t} + \beta \frac{\partial \bar{z}}{\partial x} + a U_T \frac{\partial^3 h}{\partial x^3} = 0,$$

$$(49) \qquad U_T \frac{\partial^3 \bar{z}}{\partial x^3} + k_0^2 U_T \frac{\partial \bar{z}}{\partial x} + \frac{\partial^3 h}{\partial x^2 \partial t} - k_0^2 \frac{\partial h}{\partial t} + \beta \frac{\partial h}{\partial x} + b U_T \frac{\partial^3 h}{\partial x^3} = 0$$

where

$$(50) \qquad k_0^2 = \frac{rf^2}{g}.$$

These equations may be satisfied by a trigonometric wave solution of the form: $\bar{z}, h \sim \exp i(kx + \nu t)$, provided

$$(51) \qquad \nu = \frac{\beta}{k} - \nu_b \pm \sqrt{\nu_b^2 - \nu_a^2}$$

where

$$(52) \qquad \nu_b = \frac{k_0^2 \beta + b U_T k^4}{2k(k_0^2 + k^2)},$$

$$(53) \qquad \nu_a^2 = a U_T^2 k^2 \frac{k_0^2 - k^2}{k_0^2 + k^2}.$$

When $\nu_b^2 > \nu_a^2$ we obtain stable waves with the speed of propagation $c = -\nu/k$. When $\nu_b^2 < \nu_a^2$, the waves are unstable; their speed of propagation is $c = -\beta/k^2 + \nu_b/k$, and their growth-rate $\nu_i = \sqrt{\nu_a^2 - \nu_b^2}$. For the values of a and b given in (47), the speed of propagation and the growth-rate are shown in fig. 6 as functions of the wave length L and the thermal wind U_T. The quantities L, ν_i, c and U_T are in the diagrams represented as the non-dimensional quantities $k_0 L/2\pi$, ν_i/ν_0, c/U_0, and U_T/U_0, respectively, where

$$(54) \qquad \nu_0 = \frac{\beta}{k_0}, \qquad U_0 = \frac{\beta}{k_0^2}.$$

This has the advantage that the diagrams are valid for all latitudes and for all values of the coefficient r in (50), which involves the static stability. In latitude 45°, and with the value of r given in (47), we find

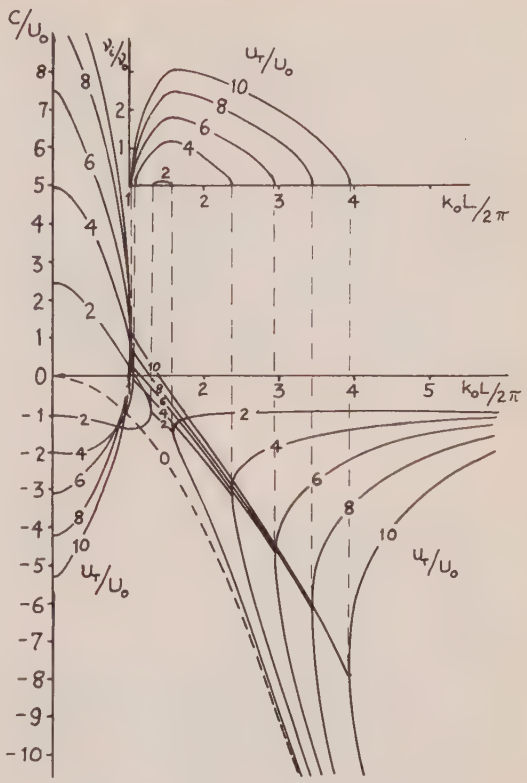


Fig. 6. Upper part: Growth rate ν_i/ν_0 of unstable waves as a function of wave-length $K_0 L/2\pi$, for different values of thermal wind U_T/U_0.
Lower part: Speed of propagation c/U_0 of stable and unstable waves as a function of wave-length $K_0 L/2\pi$, for different values of thermal wind U_T/U_0.
Dashed curve: speed of propagation of Rossby-waves ($U_T = 0$).

$$\frac{2\pi}{k_0} = 2{,}200 \text{ km}, \qquad \nu_0 = (2 \text{ days})^{-1},$$

$$U_0 = 2.0 \text{ m/sec}.$$

We notice that the waves are unstable in an intermediate band of wave lengths, in agreement with the findings of FJØRTOFT (1950) and PHILLIPS (1951). The growth-rate is a maximum when $k_0 L/2\pi \approx 1.5$, which with the numerical values given above corresponds to a wave length of about 3,300 km. The wave length of maximum instability is seen to increase in proportion to the square root of the static stability, and will also vary in inverse proportion to the Coriolis parameter; this is in good agreement with synoptic ex-

perience. Thus, for less stable air, or in higher latitudes, we will obtain considerably smaller values of the wave length of maximum instability. The values of the growth-rate and the speed of propagation obtained from the diagrams seem to be reasonable.

Thus the results obtained from the two-parametric model are plausible, as far as the behaviour of wave perturbations are concerned. This gives us some reason to hope that this model will give good results also when the equations (40), (41) are applied in non-linear form. Should these equations prove insufficient, there is still the possibility of increasing the number of parameters in the model.

Acknowledgement

The author wishes to express his thanks to Professor C. G. Rossby for his encouragement and for many valuable discussions, and to the staff members and visitors at the Institute of Meteorology, University of Stockholm, in particular to Dr. Fjørtoft and Dr. Eady, for many helpful discussions.

REFERENCES

Berson, F. A., 1950: On the factors controlling the instability of long waves in zonal currents. *Arkiv för Geofysik*, Stockholm. **1,** No. 9.

Bjerknes, J., and Holmboe, J., 1944: On the theory of cyclones. J. Meteor. **1,** No. 1.

Bjerknes, J., 1951: Extratropical cyclones. *Compendium of Meteorology*, A. M. S.

Bolin, B., and Charney, J. G., 1951: Numerical tendency computations from the barotropic vorticity equation. *Tellus*, **3,** No. 4.

Charney, J. G., 1947: The dynamics of long waves in a baroclinic westerly current. *J. Meteor*, **4,** No. 5.

— 1948: On the scale of atmospheric motions. *Geof. Publ.* **17,** No. 2.

— 1949: On a physical basis for numerical prediction of large-scale motions in the atmosphere. *J. Meteor.* **6,** No. 6.

— 1951: Dynamic forecasting by numerical process. *Compendium of Meteorology*, A. M. S.

Charney, J. G., and Eliassen, A., 1949: A numerical method for predicting the perturbations of the middle latitude westerlies. *Tellus* **1,** No. 2

Charney, J. G., Fjørtoft, R., von Neumann, J., 1950: Numerical integration of the barotropic vorticity equation. *Tellus*, **2,** No. 4.

Craig, R. A., 1945: A solution of the non-linear vorticity equation for atmospheric motion. *J. Meteor.* **2,** No. 3.

Eady, E. T., 1949: Long waves and cyclone waves. *Tellus*, **1,** No. 3.

Eliassen, A., 1949: The quasi-static equations of motion with pressure as independent variable. *Geof. Publ.* **17,** No. 3.

Fjørtoft, R., 1950: Application of integral theorems in deriving criteria of stability for laminar flows and for the baroclinic circular vortex. *Geof. Publ.* **17,** No. 6.

— 1951: Stability properties of large-scale atmospheric disturbances *Compendium of Meteorology*, A. M. S.

Høiland, E , 1950: On horizontal motion in a rotating fluid. *Geof. Publ.* **17,** No. 10.

Phillips, N., 1951: A simple three-dimensional model for the study of large-scale extratropical flow patterns. *J. Meteor.* **8,** No. 6.

Rossby, C. G., and collaborators, 1939: Relation between the intensity of the zonal circulation of the atmosphere and the displacements of the semipermanent centers of action. *J. Marine Res.* **2,** No. 1.

Rossby, C. G., 1942: Kinematic and hydrostatic properties of certain long waves in the westerlies. *University of Chicago, Dept. of Meteorology, Misc. Reports* No. 5.

Staff Members, The Institute of Meteorology, University of Stockholm, 1952: Preliminary report on the prognostic value of barotropic models in the forecasting of 500 mb height changes. *Tellus*, **4,** No. 1.

Sutcliffe, R. C., 1951: The quasi-geostrophic advective wave in a baroclinic zonal current. *Quart. J. R. Meteor. Soc.* **77,** p. 226.

Note on Weather Computing and the so-called $2\frac{1}{2}$-dimensional Model

By E. T. EADY
Imperial College, London

(Manuscript received 21 July 1952)

Abstract

The problem of setting up a system of equations by means of which the future large-scale flow pattern might be computed, using an electronic calculating machine, in a reasonable time and with maximum accuracy and efficiency, is discussed. Equations determining the development of a two-parameter, two-dimensional 'model', capable of representing the major features of motion of a baroclinic atmosphere, are derived. Tests of the model are made on some representative problems, the accurate three-dimensional solutions of which are known.

Now that actual computation of certain aspects of the weather forecast has become a practical possibility, interest in certain simplified "models" of atmospheric motion has increased. It seems at first paradoxical that this should be so. For although one necessary prerequisite for weather computing, an adequate network of upper-air data, has only recently become available, the main cause of optimism regarding the possibility of actually carrying out the necessary calculations in the time available has been the development of large-memory high-speed electronic computing machines. These machines work so *very* much faster than human computers that it is natural to suppose that at last we can forget about crude "models" and compute changes in the "actual" atmosphere. Practical experience in attempting to design a computation scheme is disillusioning. It is not merely that certain approximations (equivalent to replacing the "actual" atmosphere by a "model" but resulting in very little change in those aspects of the motion related to *weather*) are highly desirable from the point of view of *computational* simplicity, speed and stability. The main difficulty, which appears to result as the combined effect of 4 dimensions (3 space and 1 time) and non-linear equations, is the very rapid increase in computation time with increase in representation of "detail".

It is worth while to attempt to express the problem in general terms. At a given instant the weather situation may be regarded as represented by certain values of a number of parameters which may, but *need not*, be the values of pressure, temperature etc. at points on a 3-dimensional grid. Just as the definition on a television screen increases with increase in the number of spots so the accuracy with which the weather situation may be represented increases with increase in the number of parameters. The equations and boundary conditions of motion together with the ther-

mal and continuity equations determine the rate of change of these parameters and therefore the forecast situation, but since the equations involve details either not represented or inaccurately represented by the parameters some approximations are inevitable. The forecast will therefore contain not only errors of detail due to the use of a finite number of parameters but these parameters themselves will *necessarily* be to some extent in error. It is important to recognize these two kinds of error which are unavoidable even when the *computations* are made with complete accuracy. To distinguish them from a third type of error to be discussed later they will be referred to as "physical" or "model" errors. They arise simply and solely because we *cannot* include all the accurate data. Whether or not such errors are serious depends partly on the number of parameters, partly on the manner in which they are chosen. If the parameters are judiciously chosen it *may* be possible to represent with sufficient accuracy both the features in which we are most interested *and* those primarily responsible for changes in them, with only a comparatively small number of parameters. The problem is to discover the most efficient representation, that which enables us to calculate the important features as directly as possible, short-circuiting irrelevant detail. Ideally any significant variation in the initial values of any one or more of the parameters should correspond to a significant difference in the forecast. An advantage of looking at the problem from this point of view is that it suggests the right sense of proportion. Clearly there is no point in computing with more independent well-defined parameters than are determined by the initial data. On the other hand the forecast will contain the same number of parameters and therefore the same amount of detail: it will be as good as one obtained by more elaborate, time-wasting methods.

This abstract, mathematical approach gives us the general idea and is a good practical guide in the later stages of design of a computing scheme. In the early stages, however, the guidance is too vague and a different, physical approach which, though basically equivalent, is more specific, is preferable. Let us commence with an example. It is well known that if we are interested in motion on a grand scale we may usefully replace the "actual" 3-dimensional atmosphere by a 2-dimensional "model", the "barotropic model". There is a more or less close relation between the horizontal component of the grand-scale motion at about 500 mb in the atmosphere and the motion computed for the model. Here the model is a physically possible hydrodynamic system different from the atmosphere but behaving, in many important respects, in a similar manner. We might have set up this model directly from physical reasoning. On the other hand if we had integrated the vorticity equation along the vertical and then neglected certain terms we should have obtained the same final equations and it might be regarded as an accident that these equations correspond to a simple hydrodynamical system. From the point of view of our earlier abstract approach we should have suppressed all the parameters representing variation in the vertical. Setting up a model is equivalent to making certain approximations and since all computers have to make approximations there is no inherent defect in the use of models. On the contrary, if approximations correspond to setting up a model we can be sure that the former are at least self-consistent. It does not follow, however, that the most suitable approximations must correspond to a physically possible system. The representation to be described does not accurately correspond to any physical model though it does approximately do so. If the word "model" had not already been used rather vaguely it might be preferable to use some other word but since one advantage of the representation is the simplicity of physical interpretation and since the epithet "2 ½-dimensional" is already picturesque it is convenient to call this representation a "model". It represents an improvement on the barotropic model in so far as it contains a very crude representation of variation in the vertical, so crude as to be considered, playfully, as worth only half a dimension!

Such a crude representation, with only two parameters along each vertical, may be given some preliminary justification. In the first place the barotropic model, with only one parameter along the vertical, has already had some success. Theoretical reasoning also suggests that we may have a good sense of proportion if we use many more parameters

to describe horizontal variation than vertical variation. For, making much less stringent approximations, we may set up the equation for a *3-dimensional* model in the form: (EADY, 1949, p. 47)

$$0 = \frac{d}{dt}\left[\frac{\partial^2 p}{\partial x^2} + \frac{\partial^2 p}{\partial y^2} + \frac{\partial^2 p}{\partial (\sqrt{b}\, z)^2}\right] + \frac{df}{dy}\frac{\partial p}{\partial x}:$$

$$\sqrt{b} = \frac{\sqrt{g\frac{\partial \varphi}{\partial z}}}{f} \quad \ldots\ldots\ldots \quad (1)$$

Here p is the pressure deviation from a standard value, x and y are horizontal and z is the vertical co-ordinate. $\frac{\partial \varphi}{\partial z}$ $\left(\text{or } \frac{1}{\Theta}\frac{\partial \Theta}{\partial z}, \text{ where } \Theta \text{ is potential temperature}\right)$ is the static stability, g the acceleration of gravity, f the coriolis factor. Thus $\sqrt{b}$ is a pure number. A typical value in middle latitudes would be $\sqrt{b} \simeq 100$. Now although (1) is inaccurate for disturbances which are too intense, the form of the equation strongly suggests that $\sqrt{b}$ is a scale factor, representing the ratio of dynamically equivalent distances in the horizontal and vertical respectively. Thus the distance from ground to tropopause (say 10 km) is dynamically equivalent to a horizontal distance of about 1,000 km. If therefore we devote 2 parameters to representation along the vertical we should use 1 parameter in about 500 km in each horizontal direction. Since we need to consider the field over many thousands of kilometers in the horizontal it is clear that most of our parameters should describe horizontal variation. Unless the data are sufficiently detailed and accurate to determine with some accuracy more than one parameter per 500 km × 500 km square we are not justified in attempting to use a more accurate model. Although we cannot say without more investigation just how many parameters at a given (upper) level are well-determined by the data this number cannot be greater than the number of radio-sondes. It may, owing to inaccuracy and bad spacing of the sondes be much less. Hence it would appear that models of the kind to be described are not much, if at all, cruder than the best justified by the available data. Any attempt to represent much more detail would be a waste of time in *weather computing*. (This is not true of course for calculations designed to give theoretical insight.)

Instead of looking at the problem from the point of view of the amount of detail justified by the initial data we may look at it from the point of view of the computation time. Suppose we have an^2 parameters where $a=1$ for the barotropic model and $a=2$ for the present model. n is the number of parameters representing variation in each horizontal direction. Now with linear equations it is frequently possible to choose the parameters in such a manner that the variation of each is independent of all the others. With non-linear equations it is, however, not usually possible to separate the variations in this manner and in general the variation of each parameter depends on *all* the other parameters. It is evident that for accurate calculations the number of arithemetical operations will increase much faster than n^2. It looks as if the number of operations should vary as n^4 (as it does when the relevant differential equations are solved by means of a Green's function) but it is possible with both kinds of model so to arrange the work that the number of operations varies only as n^3. Nevertheless the time required for the computation of time-variations, and therefore for each small time-step in preparing the forecast, increases rapidly as more detail is included. The increase in time taken to calculate the complete forecast is still larger because with more detail smaller time-steps have to be taken. Without going into details it is evident that we must be content with quite moderate values of n and it is important that we should choose an efficient parametric representation. To give an idea of the magnitudes involved it may be mentioned that the writer estimates that, using the most rapid method known to him the value $n=13$ is about the largest that (corresponding to $a=2$) a high-speed electronic computer could cope with in order to prepare a forecast within a few hours. It is important to note that this is based on the assumption that the method must not lead to serious errors in the forecast parameters due to *computational* faults. Computational methods and errors associated with them will not be discussed here as they will form the subject of a sub-

sequent paper. It will suffice to note that most of the methods at present in use suffer from very serious computational errors: the number of parameters whose variation, over a reasonably long time interval, is reasonably accurately computed (as compared with the true variation of the *model*) is much smaller than the number of grid points and often very small indeed. Thus whatever method we use the amount of true information is not very large and it is in fact actually harmful to try to include more detail than we can cope with.

In barotropic motion (or, more generally, in motion without horizontal temperature gradients) there is no variation of motion with height so that only two space dimensions are involved. In baroclinic motion not only is there vertical motion but even the horizontal motion is a function of three dimensions. However, there is considerable synoptic evidence that for large-scale motion the major variations in the horizontal motion are described by supposing the thermal gradient independent of height. Hence we can construct a model of the horizontal part of the motion with two 2-dimensional fields, the field of the mean motion (averaged with regard to height) and the field of relative motion, the mean thermal wind (also averaged with regard to height). Alternatively by adding and subtracting a suitable multiple (either constant or at least independent of the fields of motion) of the relative motion to the mean motion we can express the model in terms of the motion at two representative levels. Now in horizontal barotropic motion the motion is (to a very close approximation) non-divergent so that in this case the mean motion (here the actual motion at any level) may be represented by a (scalar) stream function Ψ. In the case of the model with both mean and relative motion we shall find that each of these may be regarded as approximately non-divergent so that now in addition to the stream function for the mean-motion, Ψ, we have the stream-function for the relative motion, Φ. These will be regarded as (at a given time) independent, though of course their changes are interrelated. The idea of using a model of this type appears to be due to SUTCLIFFE (1947) who has described one in which there are two representative levels. The method of derivation of the equations given below is in some respects rather similar to Sutcliffe's though there are important differences in the way approximations are made, in the ancilliary assumptions and in the presentation of the results. The most important difference is the inclusion of the effect of vertical motion on the Ψ-field.

The definition of the Ψ and Φ fields serves merely to describe the motion succintly. In order to obtain a working model we have to derive sufficient (i. e. two) partial differential equations, with boundary conditions, which determine $\frac{\partial \Psi}{\partial t}, \frac{\partial \Phi}{\partial t}$ in terms of Ψ and Φ only, (or in terms of quantities which can be computed when Ψ and Φ are given). In order to do this it will be necessary to make further assumptions or postulates some of which may be regarded as fairly plausible, others somewhat arbitrary. The final test of these assumptions is the closeness with which the model simulates the behaviour of the atmosphere and for this reason a "test" on a problem which has been solved in three-dimensions, with much less stringent assumptions, will be included. Some of the assumptions made are not absolutely necessary but have the advantage of simplifying the presentation. By making more complicated assumptions it may be possible to improve the fidelity of the model. Alternatively a similar result may be obtainable if suitable (empirically determined) weighting factors are included. Actual use of the model will indicate what type of modification is most effective.

We shall suppose that the model represents motion on a large scale and that this motion is quasi-geostrophic, that is to say that at all or almost all points the geostrophic formula gives a fairly good approximation to the velocity field and also the curl of the velocity field. The horizontal divergence cannot of course be computed directly but is given with fair accuracy by the vorticity equation, obtained by eliminating the pressure field from the equations of motion. When motion is on a large scale and the Richardson number large compared with unity we may probably ignore the contribution to vorticity change due to overturning in a vertical plane and write:

$$\mathrm{div}_H v\,(f + curl_H v) + \frac{d}{dt}(f + curl_H v) = 0 \quad (2)$$

If at the same time we use (in place of the "actual" atmosphere) an incompressible model with the same static stability (but little variation of mean density with height) then we have for the continuity equation

$$\mathrm{div}_H\, v + \frac{\partial v_z}{\partial z} = 0 \ldots\ldots\ldots (3)$$

whence:

$$(f + curl_H\, v)\frac{\partial v_z}{\partial z} = \frac{d}{dt}(f + curl_H\, v) \ldots (4)$$

As is well known, a much more accurate continuity equation, of similar form to (3) is obtained when instead of z we use for a vertical coordinate the mean pressure p and at the same time replace v_z by $v_p = \frac{dp}{dt}$. The only modification in (4) is the replacement of $\frac{\partial v_z}{\partial z}$ by $\frac{\partial v_p}{\partial p}$ and in fact the whole of the subsequent analysis may be carried out using a "pressure" vertical coordinate. Then instead of making our approximations in the continuity equation we make them in the "thermal wind" equation. The results are similar in form and the only difference is in physical interpretation. In the present (spatial) model the means are to be interpreted as height-means with equal weighting for equal height difference. In the alternative (pressure) model they are to be interpreted as pressure-means. It may be that something between these extremes gives best results in practice or alternatively suitable weighting factors (relative to either interpretation) may be developed empirically. In the subsequent analysis it will be assumed that we are using a spatial vertical coordinate.

Another approximation will now be made. In equation (4) we shall suppose that in most regions $|curl_H\, v|$ may be neglected in comparison with f so that:

$$f\frac{\partial v_z}{\partial z} = \frac{d}{dt}(f + curl_H\, v) \ldots\ldots (5)$$

This is certainly not true in the vicinity of a "jet" and the relative error may also be large near intense depressions or anticyclones. In general the errors are larger at high than at low levels. However, for motion on a *large* scale we may expect to find the approximation not too bad in most regions. From the practical point of view the seriousness of the error depends on the extent to which it can be compensated for by the introduction of (readily computed) "weighting factors". Here we may note that the error may be regarded as associated with incorrect "weighting" of $\frac{\partial v_z}{\partial z}$ in respect to height. Now we have seen that the appropriate weighting of other quantities is to some extent uncertain due to the incompatible requirements for simplicity of the thermal wind and continuity relationships. Moreover we shall subsequently need to adopt a rather crude and schematic distribution for v_z since our data are insufficient to determine its true variation. The total effect of all such errors cannot easily be estimated a priori. Hence the method will be to develop the simplest possible system of equations as a basis for experiment and subsequent critical re-examination.

The stream functions Ψ and Φ are defined as mean values over a certain depth of atmosphere, the boundaries of which will be taken to be $+z_0$, $-z_0$. Thus $-z_0$ corresponds to the surface of the earth and $+z_0$ the effective "top" of the atmosphere. In fact, of course, there is no top to the atmosphere but there is some synoptic and also theoretical evidence (based on perturbation theory — cf. EADY 1949) that large scale disturbances are mainly confined to the troposphere and extend only a short distance into the stratosphere. Hence the effective depth $2z_0$ will be supposed to correspond to a little more than the height of the tropopause but once again some modification may be desirable in conjunction with the use of weighting factors. z_0 is not a constant in fact but unless a simple approximate relation to the Ψ and Φ fields can be found it may be necessary to use a mean value independent of the field of motion. This mean value certainly decreases slowly with latitude but the exact effect of this variation will not be included in the present analysis. The effective depth also varies because of unevenness of the earth's surface and it is known that this is responsible for the existence of orographic stationary long waves. Since this particular variation of z_0 is well defined beforehand it is easy to include its

effects in the equations but for simplicity this modification will be omitted in the present treatment. Then if we suppose both the base and "top" of the effective part of the atmosphere as substantially flat we may take the boundary conditions to be $v_z = 0$ at $z = \pm z_0$. Integrating (5) with respect to z between these limits:

$$0 = \int_{-z_0}^{+z_0} \frac{d}{dt}(f + curl_H v)\, dz \quad \ldots\ldots (6)$$

By definition the velocity field in the model at any level z is:

$$\left.\begin{aligned} v_x &= -\frac{\partial \Psi}{\partial y} - \frac{z}{z_0}\frac{\partial \Phi}{\partial y} \\ v_y &= \frac{\partial \Psi}{\partial x} + \frac{z}{z_0}\frac{\partial \Phi}{\partial x} \end{aligned}\right\} \ldots\ldots (7)$$

approximately. This field of motion is strictly non-divergent but a good enough approximation for all purposes except the *direct* calculation of divergence. In high latitudes Ψ and Φ are very nearly constant multiples of the mean pressure and temperature fields respectively, consistent with the geostrophic approximation and the relatively small value of $\frac{df}{dy}$.

In low latitudes Ψ and Φ are better determined directly from the wind data. Some slight adjustments will in practice be needed to fit the data in middle latitudes and obtain the best Ψ and Φ representation.
Writing:

$$\begin{aligned} v_{x_0} &= -\frac{\partial \Psi}{\partial y}; \quad v_{y_0} = \frac{\partial \Psi}{\partial x}; \\ v_{x_1} &= -\frac{\partial \Phi}{\partial y}; \quad v_{y_1} = \frac{\partial \Phi}{\partial x} \quad \ldots\ldots (8) \end{aligned}$$

and also:

$$\zeta_0 = curl_H v_0 = \nabla_H^2 \Psi; \ \zeta_1 = curl_H v_1 = \nabla_H^2 \Phi \quad (9)$$

we have on substituting in (6):

$$0 = \int_{-z_0}^{+z_0} \left\{ \left(v_{y_0} + \frac{z}{z_0} v_{y_1}\right)\frac{df}{dy} + \left[\left(v_{x_0} + \frac{z}{z_0} v_{x_1}\right)\frac{\partial}{\partial x} + \right.\right.$$
$$\left.\left. + \left(v_{y_0} + \frac{z}{z_0} v_{y_1}\right)\frac{\partial}{\partial y} + \frac{\partial}{\partial t}\right]\left(\zeta_0 + \frac{z}{z_0}\zeta_1\right)\right\} dz \quad (10)$$

where the contribution from $\int_{-z_0}^{+z_0} v_z \frac{\partial \zeta}{\partial z}\, dz$ has been neglected. Perturbation theory suggests that the contribution from this term is small, compared with that of the remaining terms, when the Richardson number is large, as may be verified by substituting the true values of the velocities corresponding to the development of a baroclinic wave (see EADY 1949). We shall tentatively assume that the contribution from this term corresponding to vertical advection of vorticity may be neglected, in more general conditions, over most of the region concerned. The integrand in (10) is, so far as variation with z is concerned, a quadratic form and odd powers of z integrate to zero. Hence we obtain:

$$v_{y_0}\frac{df}{dy} + \left(v_{x_0}\frac{\partial}{\partial x} + v_{y_0}\frac{\partial}{\partial y} + \frac{\partial}{\partial t}\right)\zeta_0 +$$
$$+ \frac{1}{3}\left(v_{x_1}\frac{\partial}{\partial x} + v_{y_1}\frac{\partial}{\partial y}\right)\zeta_1 = 0 \quad \ldots\ldots (11)$$

and if we write:

$$\frac{D}{Dt} \equiv \left(v_{x_0}\frac{\partial}{\partial x} + v_{y_0}\frac{\partial}{\partial y} + \frac{\partial}{\partial t}\right) \quad \ldots (12)$$

corresponding to Langrangian differentiation following the *mean* motion, then

$$\frac{D}{Dt}(f + \zeta_0) = -\frac{1}{3}\left(\frac{\partial \Phi}{\partial x}\frac{\partial \zeta_1}{\partial y} - \frac{\partial \Phi}{\partial y}\frac{\partial \zeta_1}{\partial x}\right) \quad (13)$$

The term on the right hand side may be called the *development* term. It represents the change in absolute vorticity of a vertical column moving with the mean motion (but not of course consisting always of the same air particles) and expresses the difference between baroclinic motion and the motion of an "equivalent" barotropic model. We shall use the notation:

$$\left(\frac{\partial A}{\partial x}\frac{\partial B}{\partial y} - \frac{\partial A}{\partial y}\frac{\partial B}{\partial x}\right) \equiv \frac{\partial (A, B)}{\partial (x, y)} \equiv \{A, B\} \quad (14)$$

for the Jacobian of any two quantities A, B, with respect to x, y. Then for computational purposes (13) is conveniently written:

$$-\nabla_H\left(\frac{\partial\Psi}{\partial t}\right)=\frac{df}{dy}\frac{\partial\Psi}{\partial x}+\left\{\Psi,\nabla_H^2\Psi\right\}+$$
$$+\frac{1}{3}\left\{\Phi,\nabla_H^2\Phi\right\} \quad \ldots\ldots\ldots\ldots (15)$$

This gives one of the required equations, relating $\frac{\partial\Psi}{\partial t}$ to Ψ and Φ. We now need an equation to determine $\frac{\partial\Phi}{\partial t}$. Now from the geostrophic relations:

$$\frac{\partial v_x}{\partial z}=-\frac{g}{f}\frac{\partial\varphi}{\partial y};\quad \frac{\partial v_y}{\partial z}=\frac{g}{f}\frac{\partial\varphi}{\partial x} \quad .. (16)$$

where φ is the logarithm of potential temperature, we obtain on comparison with (7):

$$\frac{\partial\Phi}{\partial x}=\frac{gz_0}{f}\frac{\partial\varphi}{\partial x};\quad \frac{\partial\Phi}{\partial y}=\frac{gz_0}{f}\frac{\partial\varphi}{\partial y} \quad .. (17)$$

If we ignore the slow variation of $\frac{z_0}{f}$ with latitude and remember that φ is a function of (x, y, z) whereas $\Phi=\Phi(x, y)$ by definition, then

$$\Phi=\frac{gz_0}{f}\cdot\bar{\varphi} \quad \ldots\ldots\ldots (18)$$

where $\bar{\varphi}$ denotes a mean with respect to z. Since φ varies also with z we shall write:

$$\varphi=\bar{\varphi}+Bz=\frac{f}{gz_0}\Phi+Bz \ldots (19)$$

where $B=\overline{\frac{\partial\varphi}{\partial z}}$ is the mean static stability. The relative variations of B are not usually very large and they will be ignored in subsequent calculations. From (19) it follows, if we ignore the slow variation of $\frac{f}{z_0}$, that:

$$\frac{d\varphi}{dt}=\frac{f}{gz_0}\frac{D\Phi}{Dt}+Bv_z \quad \ldots\ldots (20)$$

If for the present we suppose the adiabatic approximation to be sufficiently accurate we have:

$$0=\frac{f}{gz_0}\frac{D\Phi}{Dt}+Bv_z \quad \ldots\ldots (21)$$

and the change in Φ is apparently a function of z. However, our model is incapable of representing variation of this kind and all we require is an estimate of the *mean* change. Integrating between $-z_0$ and $+z_0$ we get:

$$\frac{D\Phi}{Dt}=-\frac{gB}{2f}\int_{-z_0}^{+z_0}v_z\,dz \quad \ldots\ldots (22)$$

which would be the required equation if we could express the mean value of v_z in terms of Φ and Ψ.

To do this we once again integrate equation (5) but instead of adding the contributions from each half of the atmosphere ($z=-z_0$ to $z=0$ and $z=0$ to $z=+z_0$) we subtract them. Then if $v_z(0)$ is the vertical velocity at the middle level:

$$2f\,v_z(0)=\int_{-z_0}^{0}-\int_{0}^{+z_0}\left\{\left(v_{y_0}+\frac{z}{z_0}v_{y_1}\right)\frac{df}{dy}+\right.$$
$$+\left[\left(v_{x_0}+\frac{z}{z_0}v_{x_1}\right)\frac{\partial}{\partial x}+\right.$$
$$\left.\left.+\left(v_{y_0}+\frac{z}{z_0}v_{y_1}\right)\frac{\partial}{\partial y}+\frac{\partial}{\partial t}\right]\left(\zeta_0+\frac{z}{z_0}\zeta_1\right)\right\}dz \quad (23)$$

if once again the vertical advection of vorticity is neglected. On evaluation the even powers of z integrate to zero and we obtain:

$$-\frac{2f}{z_0}v_z(0)=v_{y_1}\frac{df}{dy}+\left(v_{x_1}\frac{\partial}{\partial x}+v_{y_1}\frac{\partial}{\partial y}\right)\zeta_0+$$
$$+\left(v_{x_0}\frac{\partial}{\partial x}+v_{y_0}\frac{\partial}{\partial y}+\frac{\partial}{\partial t}\right)\zeta_1 \ldots (24)$$

There is, of course, no necessary relation between the mean value of v_z and the value of $v_z(0)$. However, with large scale disturbances we may expect to find that v_z is rather simply distributed with, in *most* regions, the same sign for all z, a simple maximum near the middle level ($z=0$) and of course zero values at $z=\pm z_0$. The simplest distribution which satisfies these requirements is the parabolic one in which $v_z\propto(z_0-z)(z_0+z)$. Then we have:

$$\frac{1}{2z_0}\int_{-z_0}^{+z_0}v_z\,dz=\frac{2}{3}v_z(0) \quad \ldots\ldots (25)$$

As a check we may note that the theoretical distribution for a growing baroclinic disturbance (EADY 1949) does not differ very greatly from this assumed distribution. The principal error arises from the fact that our model is not able to represent the (comparatively small) *phase* change of v_z with height: in this example we know that the distribution of v_z with z is not independent of x, y. However, this feature is associated with the fact that the disturbance is rapidly intensifying and it may be less marked in average conditions. We may note that an error of the same kind is involved in the assumption of constant thermal wind: in a growing disturbance there is a (comparatively small) phase change with height in the φ-field.

From (25) and (24) we obtain an expression for the mean vertical velocity. Substitution in (22) gives:

$$\frac{D\Phi}{Dt} = \frac{1}{3}\frac{gB}{f^2}\cdot z_0^2\left[v_{y_1}\frac{df}{dy} + \left(v_{x_1}\frac{\partial}{\partial x} + v_{y_1}\frac{\partial}{\partial y}\right)\zeta_0 + \right.$$
$$\left. + \left(v_{x_0}\frac{\partial}{\partial x} + v_{y_0}\frac{\partial}{\partial y} + \frac{\partial}{\partial t}\right)\zeta_1\right] \ldots (26)$$

If, as in equation (1), we write $b = \frac{gB}{f^2}$, which as we have seen is a pure number to be interpreted as the square of the horizontal-vertical scale factor, then (26) may be written:

$$\frac{3}{bz_0^2}\frac{D\Phi}{Dt} = \frac{D}{Dt}\nabla_H^2\Phi + \frac{df}{dy}\frac{\partial\Phi}{\partial x} + \left\{\Phi, \nabla_H^2\Psi\right\} \quad (27)$$

where the notation of (14) has been used for the Jacobian. For computational purposes this result is conveniently rewritten:

$$\left(\frac{3}{bz_0^2} - \nabla_H^2\right)\left(\frac{\partial\Phi}{\partial t}\right) = \frac{df}{dy}\frac{\partial\Phi}{\partial x} + \left\{\Phi, \nabla_H^2\Psi\right\} +$$
$$+ \left\{\Psi, \nabla_H^2\Phi\right\} - \frac{3}{bz_0^2}\left\{\Psi, \Phi\right\} \ldots (28)$$

We now have a pair of equations (15) and (28) for determing $\frac{\partial\Psi}{\partial t}$ and $\frac{\partial\Phi}{\partial t}$. The right-hand sides of these equations are known functions so that to determine $\frac{\partial\Psi}{\partial t}$ we have a Poisson differential equation while $\frac{\partial\Phi}{\partial t}$ is determined by a Helmholtz equation. The sign of the constant $\frac{3}{bz_0^2}$ in the Helmholtz equation is such as to ensure a unique readily computed solution. To solve the equations we need to apply suitable boundary conditions. If Ψ and Φ are given over the whole of the earth's surface this is a simple matter. The "boundary" condition is that $\frac{\partial\Psi}{\partial t}$ and $\frac{\partial\Phi}{\partial t}$ must not have singularities anywhere and this determines the functions uniquely. If the data are given over a hemisphere we may suppose that both functions vanish on the equator. With data over more limited regions the appropriate boundary conditions are less obvious and this question will be discussed in a subsequent paper. We may note that the equations (15) and (28) are easily adapted to computations over a spherical surface. Whether it is necessary or convenient to do this is another question the discussion of which is postponed.

In the above account we have neglected two features which *in the long run* must play an important part in determining atmospheric motion — surface frictional drag and influx and efflux of heat through radiation and convection from the earth's surface. Formally it is a simple matter to include both these features in the model. For example if (to make a crude estimate) we suppose the surface stress proportional to and in the direction opposite to the geostrophic wind at $z = -z_0$, then the torque acting on the column of air above will be proportional to the curl of the wind field at $z = -z_0$, i.e. to $\nabla_H^2(\Psi - \Phi)$. This torque measures an additional rate of change of the vorticity of the column and we have only to add this term to the right-hand side of (15) to represent the effect of friction. The effects of heating and cooling are expressed by adding Q, the rate of change of φ due to non-adiabatic heating or cooling, to the left-hand side of (21). Proceeding in the same way as before we obtain an additional term, $\frac{3f}{Bz_0}\bar{Q}$ where $\bar{Q}$ is the mean value of Q averaged with respect to z, on the right-hand side of (28). Of course,

in order to apply this result we must have some simple means of estimating Q when Ψ and Φ are given.

Perhaps the best test of the formulae would be a prolonged series of experimental calculations for comparison with observed behaviour. Alternatively we may test the model for types of motion where the 3-dimensional solutions are known. We have already made some comparisons with the results of perturbation theory. A more elaborate test may be made by seeing to what extent some of the quantitative results of perturbation theory are reproduced by the model. Undisturbed horizontal baroclinic flow may be represented by a mean flow $U = -\left(\frac{\partial \Psi}{\partial y}\right)_0$ and a thermal wind $T = -\left(\frac{\partial \Phi}{\partial y}\right)_0$ where U and T are constants. If now we superpose a small perturbation represented by stream functions Ψ_1, and Φ_1, we obtain, on substituting in (15) and (28) and picking out the first order terms, the perturbation equations:

$$0 = \left(U\frac{\partial}{\partial x} + \frac{\partial}{\partial t}\right)\nabla_H^2 \Psi_1 + \frac{df}{dy}\frac{\partial}{\partial x}\Psi_1 + \\ + \frac{1}{3} T \frac{\partial}{\partial x}\nabla_H^2 \Phi_1 \ldots\ldots (29)$$

and:

$$0 = \left(U\frac{\partial}{\partial x} + \frac{\partial}{\partial t}\right)\left(\nabla_H^2 - \frac{3}{bz_0^2}\right)\Phi_1 + \frac{df}{dy}\frac{\partial}{\partial x}\Phi_1 + \\ + T\frac{\partial}{\partial x}\left(\nabla_H^2 + \frac{3}{bz_0^2}\right)\Psi_1 \ldots. (30)$$

Corresponding to the assumptions made in the 3-dimensional theory (EADY 1949) we suppose that $\sqrt{b}\, z_0 \equiv H$ and $\frac{df}{dy}$ may be taken as effectively constant. Then the simultaneous equations (29) and (30) evidently have solutions of the form:

$$\Psi_1 = E e^{i(\lambda x + \mu y + \vartheta t)} : \quad \Phi_1 = F e^{i(\lambda x + \mu y + \vartheta t)} \quad (31)$$

where E, F, λ, μ, ϑ are constants. Equations (29) and (30) will be satisfied if:

$$\begin{cases} (U_w - U_r)\ E + \frac{1}{3} T F = 0 \\ [U_w(1+X) - U_r]\ F + T(1-X)\ E = 0 \end{cases} (32)$$

where we have written:

$$U_w = U + \frac{\vartheta}{\lambda}; \quad U_r = \frac{1}{(\lambda^2+\mu^2)} \cdot \frac{df}{dy}; \\ X = \frac{3}{(\lambda^2+\mu^2) H^2} = \frac{3}{(\lambda^2+\mu^2) bz_0^2} \ldots (33)$$

Eliminating E/F from equations (32):

$$(1+X)\left(\frac{U_w}{U_r}\right)^2 - (2+X)\left(\frac{U_w}{U_r}\right) + 1 - \\ - \frac{T^2}{3 U_r^2}(1-X) = 0 \ldots\ldots (34)$$

whence:

$$\frac{U_w}{U_r} = \frac{(2+X) \pm \sqrt{X^2 + \frac{4}{3}\frac{T^2}{U_r^2}(1-X^2)}}{2(1+X)} \quad (35)$$

The boundary conditions in the horizontal directions require that both λ and μ should be real. Hence X and $\left(\frac{T}{U_r}\right)^2$ are necessarily real and positive. The disturbances will be unstable if, and only if, ϑ is complex, i. e. if $\frac{U_w}{U_r}$ is complex. The condition for instability is therefore:

$$X^2 + \frac{4}{3}\frac{T^2}{U_r^2}(1-X^2) < 0 \ \ldots (36)$$

Now $\sqrt{X}$ is a number proportional to the ratio of the "effective" wavelength $\frac{2\pi}{\sqrt{\lambda^2+\mu^2}}$ to the "dynamic depth" 2H. There will be instability for *some* wavelengths if (rearranging (36))

$$\frac{U_r^2}{X^2 T^2} = \left(\frac{H^2}{3}\frac{df}{dy}\right)^2 < \frac{4}{3}\frac{1}{X^2}\left(1 - \frac{1}{X^2}\right) \quad (37)$$

for any value of $(\lambda^2+\mu^2)$, i. e. for any positive X. The maximum value of the right-hand side is $1/3$ (when $X^2 = 2$). Hence the condition for instability of the initial flow is:

$$\left|\frac{H^2}{T}\frac{df}{dy}\right| < \sqrt{3} \ \ldots\ldots\ldots (38)$$

The condition may also be written in the form:

$$s \equiv \left|\frac{A}{B}\right| > \frac{1}{\sqrt{3}} \frac{z_0}{f} \frac{df}{dy} \ldots\ldots (39)$$

where s is the slope of the (unperturbed) isentropic surfaces. This result is of considerable interest, though it has yet to be checked by three-dimensional calculations. The corresponding wavelength for the disturbances which first become unstable as T is increased is given by $X=\sqrt{2}$. Eliminating H^2 we obtain

$$\frac{1}{(\lambda^2+\mu^2)} \frac{df}{dy} = \sqrt{\frac{2}{3}} \cdot T \ldots\ldots (40)$$

and this result may be compared with the stability criterion discovered by CHARNEY (1947).

The simplicity of the above calculations illustrates the power of the 2 ½-dimensional model. As a more precise test of its accuracy we shall consider the case when $\frac{df}{dy}$ is neglected. Then $U_r = 0$ and in place of (34) we have:

$$U_w^2 = \frac{T^2}{3}\left(\frac{1-X}{1+X}\right) \ldots\ldots (41)$$

In this case there always exist unstable waves, as is evident also from (38). The condition to be satisfied is that $X > 1$ or:

$$\frac{1}{\sqrt{\lambda^2+\mu^2}} > \frac{1}{\sqrt{3}} \cdot \sqrt{b}\, z_0 \simeq \frac{1}{1.732} \cdot \sqrt{b}\, z_0 \quad (42)$$

This compares with the true value:

$$\frac{1}{\sqrt{\lambda^2+\mu^2}} > \frac{1}{1.1997} \cdot \sqrt{b}\, z_0 \ldots\ldots (43)$$

(EADY 1949 p. 39). The only error is in the numerical factor and this is clearly of a kind which *could* be eliminated by modifying the constants in the equations (15) and (28).

It is easily verified that the disturbance of maximum growth rate (maximum imaginary part of ϑ) corresponds to $\mu = 0$ and maximum $\frac{U_w^2}{X}$. From (41) we find that $X = 1+\sqrt{2}$ gives the corresponding wavelength so that:

$$\frac{1}{\lambda} = \frac{\sqrt{1+\sqrt{2}}}{\sqrt{3}} \cdot \sqrt{b}\, z_0 \simeq \frac{1}{1.115} \cdot \sqrt{b}\, z_0 \quad (44)$$

as compared with the true value:

$$\frac{1}{\lambda} = \frac{1}{0.8031} \cdot \sqrt{b}\, z_0 \ldots\ldots (45)$$

The corresponding value of ϑ is purely imaginary as in the accurate calculations. The numerical magnitude deduced from (44) and (41) is:

$$|\vartheta| = (\sqrt{2}-1) \cdot \frac{T}{\sqrt{b}\, z_0} \simeq 0.4142 \frac{T}{\sqrt{b}\, z_0} \quad (46)$$

as compared with the true value:

$$|\vartheta| = 0.3098 \frac{T}{\sqrt{b}\, z_0} \ldots\ldots (47)$$

The value of $\left|\frac{\vartheta}{\lambda}\right|$ computed for the model is:

$$\left|\frac{\vartheta}{\lambda}\right| = \frac{1}{\sqrt{3}\sqrt{1+\sqrt{2}}} \cdot T \simeq 0.372\, T \ldots\ldots (48)$$

which compares favourably with the true value:

$$\left|\frac{\vartheta}{\lambda}\right| = 0.3858\, T \ldots\ldots (49)$$

The above is probably a fairly severe test of the model. For comparison we may note that the assumption that temperature is horizontally advected (neglect of vertical motion) leads to values of $|\vartheta|$ which are very much too large for the short wave-lengths — in fact there is no minimum wave-length for instability. For very long waves the effect of vertical motion is much less but on the other hand it is precisely in these conditions that an equivalent barotropic model can be constructed to reproduce the behaviour of the mean flow with some degree of accuracy.

The writer would like to express his thanks to Professor C.-G. Rossby and members of Institutet för Meteorologi, Stockholm, for lively and constructive criticism, during a recent visit, of the ideas presented here.

REFERENCES

Bolin, B., 1951: On the influence of the earth's orography on the general character of the westerlies. *Tellus*, **2**, 184—195.

Charney, J. G., 1947: The dynamics of long waves in a baroclinic westerly current. *J. Meteor.*, **5**, 135—162.

— 1949: On a physical basis for numerical prediction of large-scale motions in the atmosphere. *J. Meteor.*, **6**, 371—385.

— and Eliassen, A., 1949: A numerical method for predicting the perturbations of the middle latitude westerlies. *Tellus*, **1**, 38—54.

Charney, J. G., Fjørtoft, R., and von Neumann, J., 1950: Numerical integration of the barotropic vorticity equation. *Tellus*, **2**, 237—254.

Eady, E. T., 1949: Long waves and cyclone waves. *Tellus*, **1**, 33—52.

Fjørtoft, R., 1950: Application of integral theorems in deriving criteria of stability for laminar flows and for the baroclinic circular vortex. *Geophys. Publ.*, **17**, No. 6.

Phillips, N. A., 1951: A simple three-dimensional model for the study of large-scale extratropical flow patterns. *J. Meteor.*, **8**, 381—394.

Richardson, L. F., 1922: *Weather prediction by numerical process*. C. U. Press, Cambridge, England.

Sawyer, J. S., and Bushby, F. H., 1951: Note in the numerical integration of the equations of meteorological dynamics. *Tellus*, **3**, 201—203.

Sutcliffe, R. C., 1947: A contribution to the problem of development. *Quart. J. R. Meteor. Soc.*, **73**, 370—383.

— and Forsdyke, A. G., 1950: The theory and use of upper air thickness patterns in forecasting. *Quart. J. Roy. Met. Soc.*, **76**, 189—217.

Some Remarks on High-Speed Automatic Computers and their Use in Meteorology

By GEORGE W. PLATZMAN
University of Chicago

(Manuscript received 30 June 1952)

Abstract

The logical design of an electronic digital computer is described. A simple computation problem is proposed; this is used to illustrate the main steps leading to formulation of a problem for machine computation — the propram and the flow diagram. In conclusion, some general comments are made concerning areas of application of high-speed computers in meteorology.

High-speed digital computers now are to some extent available to meteorologists for research. Doubtless there will be intensive developments in certain fields of theoretical and applied meteorology where heretofore progress has been inhibited by the prodigous amounts of computation required to yield a useful result. As this work advances, it is likely to take on some of the aspects of an engineering science, and the underlying principles of the use of computers for meteorological problems may tend to become obscured by a superstructure of special methods of computation and procedure.

It is inevitable that high-speed computers will open new avenues of investigation in theoretical studies of the atmosphere, and although among meteorologists there is widespread latent interest in these prospects, it is probably too early to predict when to expect results of practical value — say, to the forecaster. Nevertheless, it seems rather evident that to maintain a well-balanced approach to problems that offer some promise of yielding to computation methods, there should be effective communication between those who gain experience in the techniques of computation, and those whose contacts are mainly with applied meteorological problems — for example, in the fields of forecasting or synoptic meteorology.

With this thought in mind, the writer has undertaken to sketch a broad view of some of the steps that enter into the task of domesticating a high-speed computer for meteorological purposes, and although the view is from a very limited level of experience in these matters, it may nevertheless help to illuminate for the meteorologist this novel area of research and to implement his interests in a field that he may sense is intriguing but which otherwise may seem obscure. The results of this endeavor are reported in what follows; they are grouped under four headings: (1) a computer, (2) a problem, (3) a computation, (4) concluding remarks.

The substance of this discussion has been taken from the reports by Goldstine and von Neumann, to which reference is made at the end of the paper. The reader should consult these reports for a more systematic and thorough treatment of the logical design of a high-speed computer and a more detailed exposition of methods of preparing problems for computation.

1. A computer

An electronic digital computer is an extremely complex instrument containing many thousands of components and involving application of the most advanced techniques of electronic engineering. Fortunately, it is possible to comprehend the logical structure of the machine without understanding the design and construction whereby this logical structure is achieved, just as it is possible to know the logical structure of an automobile without understanding the construction of the various components or even the manner in which they accomplish the end result. Indeed, this is not surprising because in point of fact, the logical structure of the computer must be formulated more or less independently of the specific engineering techniques by means of which its physical counterparts ultimately are realized.

The basic logical elements of an electronic digital computer are (i) the *arithmetic* organ, (ii) the *memory*, (iii) the *control*. In addition, there must be provision for transfering data into and out of the machine — namely, an *input-output* device. The significance of the primary logical elements can be illustrated by a simple example. Let a and b stand for two numbers which, at a certain stage in a particular computation, must be combined arithmetically to yield a derived number, $F(a, b)$; for example, $F(a, b) = a+b$, or $a-b$, or $a \times b$, etc. Initially, a and b are stored in the memory; the control arranges their transfer to the arithmetic organ, instructs this organ to produce $F(a, b)$, and returns this derived number to the memory (see figure 1). (Certain operations involve only a single number; for example, $F(a) = -a$, or $F(a) = |a|$, or $F(a) =$ the first four digits of a, and so on.)

The arithmetic organ contains a small number of *registers* — analogous to the "dial" registers of an ordinary desk calculator — each of which can hold one number (that is, a series of digits) at any instant. Not many registers are required if, as in most computers, the arithmetic organ performs only one operation at a time. When a number held in a register is no longer needed, the control can erase it, or send it to a specific location in the memory. If the specified memory location already holds a number, this number is erased and the number sent from the register takes its place.

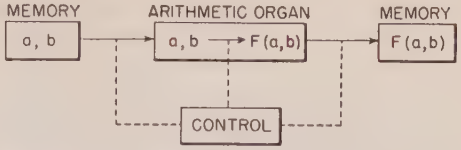


Figure 1. The basic elements and their functions.

The usefulness of a high-speed computer depends to a large extent on the *versatility of the control*, and the *capacity of the memory*. Every computation of sufficient scope to justify the use of a high-speed computer will involve a great many intermediate numbers that do not represent the final result but are produced at some stage of the computation and set aside to be used at a later stage. Furthermore, in some problems the initial data as well as the final computed result may in themselves involve a great many numbers. If the machine is to function as a *high-speed* computer, it is clearly essential to limit sharply the necessity for stopping the machine and supplying or removing numbers before the computation is finished; indeed, ideally, the machine should not be interrupted at all during the course of the computation. To provide for a completely automatic sequence of computations, there must be a fully automatic control that can process the necessary instructions (orders) without interruption, and a large-capacity memory that can accommodate simultaneously a great many numbers.

One of the simplest types of memory to visualize is built around a specially designed electron tube, the so-called "electrostatic storage tube", which in construction and operation is similar to a small television tube (iconoscope). At one end there is a network or mosaic of electrostatically charged "spots". Each of these small charged areas can be represented schematically in the manner shown

Figure 2. Schematic illustration of the networks of charged points in five identical electrostatic-storage memory tubes. The circles enclose *corresponding* points.

in figure 2. Tube 1 is pictured as a network of 16 points, each point representing a small area of electrostatic charge; tube 2 is identical, and so on. A *number* (series of digits) can be represented as equivalent to a group of charges located at *corresponding* points in the series of networks; for example, the circled point in tube 1 holds the first digit of the number, the corresponding point (circled) in tube 2 holds the second digit, and so on. The arrangement illustrated in figure 2 would in this way accommodate 16 five-digit numbers (16 points in each tube, and 5 tubes). In practice as many as 40 tubes are used, each with as many as 1,024 points (a 32×32 network instead of the 4×4 pictured above); a memory of this size would have a capacity of $32 \times 32 =$ 1,024 forty-digit numbers.

In each tube is a device for producing a narrow beam of electrons that impinges on the network of changed points, and in fact can "scan" this network point by point and row by row. The required deflections of the electron beam are produced by a properly controlled magnetic field, which means that the scanning process can be performed with extreme rapidity. The tubes are synchronized so that at any instant their respective beams impinge on corresponding points of the charged networks; in this way the various digits in a given number can be "sensed" simultaneously and transferred to a register. By proper control of the intensity of the beam, it is possible to arrange for the scanning process to *read*, or to *write*, which means in effect that a number can be transferred *from* the memory *to* the register, or *vice versa*.

The computations required to solve a given problem can be formulated as a list of *orders* that enumerate step by step, every single arithmetic manipulation that must be performed, including all necessary transfers of numbers to and from the memory. The orders are then translated into numbers according to a definite code, and as such are themselves stored in the memory (thereby establishing the link between human control and machine control of the computation). In this sense the memory holds *logical* information as well as strictly *numerical* information; both types of information are expressed as series of digits). Since the orders are now located in the memory, they can be scanned and transferred, so to speak, to the control organ, which arranges for these orders to be performed, one after the other, as they are read.

Virtually every problem of a scope that justifies the use of a high-speed computer will demand a highly versatile control organ. For example, a commonly recurring situation at various stages of the computation requires repetition (iteration) of the same sequence of operations and hence of the same sequence of orders. To illustrate: suppose we have, say, fifty pairs of numbers and that the same set of operations must be performed on each of these number pairs. Obviously, in writing out the orders we cannot afford, as a rule, to repeat the same sequence over and over again the desired number of times because the memory space available for these orders probably would be quickly exhausted. It should be sufficient to list this sequence of orders just once; so the control should be able to "go back" to the beginning of the sequence and repeat as many times as required. In fact, we should generalize this requirement in the following way: normally, the control will take up the orders in sequence as they are entered in the memory; however, at any stage by means of an order that "transfers" the control, it will proceed not to the next order but to any other designated order, and there continue in sequence.

There is another, vital feature of versatility of the control. Where a transfer of the control is required, we should in principle specify beforehand precisely the order to which the control should be transferred; or, if a sequence of orders must be iterated, we should specify beforehand exactly how many iterations must be performed. There are several difficulties here, connected with the fact that very frequently this type of information is not known until the computation has proceeded up to a certain point. If we are to avoid interrupting the computation, such decisions (to transfer

the control) must be made *internally* on the basis of specific orders provided for this purpose. Indeed, even when for example we know in advance exactly how many iterations of a given sequence of orders are needed, there would be no convenient way to terminate these iterations unless a decision can be made internally. All "decisions" of the type required can be reduced to an order providing for *conditional* transfer of control, depending on the sign (plus or minus) of a number held in one of the registers. Suppose, for example, that the first iteration of a sequence of orders produces a number a_1, the next iteration a slightly smaller number a_2, the third a still smaller number a_3, and so on, and we want to iterate until a number, say a_n, is produced that is smaller than a specific preassigned number b. This is accomplished as follows. After each iteration there is an order placing the difference $a_n - b$ in one of the registers, followed by the conditional transfer order. If this difference is positive (meaning a_n is still greater than b) the control will transfer back to the iteration; if it is negative the control will proceed *out* of the iteration to whatever order may follow in sequence. This provision for conditional transfer of control is used repeatedly in most problems. It provides for extremely versatile and fully automatic control of the computation.

Another typical application of the conditional transfer order is this: suppose we know in advance exactly the number of iterations required for a given sequence of orders — say, 30 iterations. Before the first run through the sequence the number 1 is ordered into an empty location in the memory, and just prior to each repetition of the sequence this "counting index" or "iteration index" is advanced by adding 1 to the number already there, so that at the end of each sequence this particular memory location will hold a number exactly equal to the number of times the sequence has been completed up to that stage. Further, at the end of each sequence the difference "29 minus the index" is ordered into a register and then the conditional transfer order follows. When this difference is *positive* (or zero), meaning that 30 iterations (the desired number) have not yet been completed, the control is transferred back to the beginning of the sequence; but after the 30th iteration this difference is *negative*, so the conditional transfer order carries the control out of the iteration and on to the next part of the computation.

We have now discussed the main features of the logical framework of a computer; they are (apart from an input-output device):

(i) An arithmetic organ, consisting of a few registers each to hold a single number (series of digits) at any instant. The arithmetic organ is responsible for performing basic arithmetic operations such as addition, multiplication, division.

(ii) A memory, capable of storing a great many numbers, with provisions for reading and altering its contents as the computation proceeds.

(iii) A versatile control, which responds to a list of orders stored in the memory.

A final comment regarding the *speed* of computation: in an electronic computer all internal operations are purely electronic and therefore extremely rapid. Indeed, the unit of time most convenient in this connection is the *microsecond* — one millionth of a second. A good approximation to the duration of most types of computation can be obtained by allowing about 500 microseconds for each multiplication (or division), and about 65 microseconds for each order (including multiplications and divisions). For example, in a typical meteorological calculation (the 24-hour barotropic forecast) carried out on the Princeton computer, 1,960,000 multiplications and a total of 42,300,000 orders are required. The approximate duration of this computation is therefore $500 \times 1{,}960{,}000 + 65 \times 42{,}300{,}000$ microseconds, or about 62 minutes. This makes it quite clear why an extremely involved and lengthy computation — hopelessly impractical to attempt by hand — may be quite practical for the high-speed computer.

2. A problem

The problem that will now be described has been chosen mainly to illustrate some typical steps leading to formulation of a computation for the high-speed computer. It is in fact a simple problem and requires only a rather small amount of computation, so it is not representative of the type of problems for which a high-speed computer is best

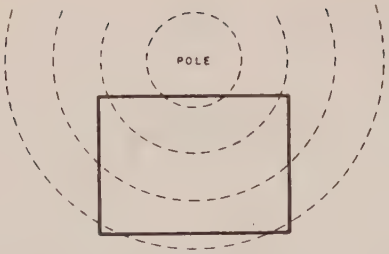


Figure 3. Polar stereographic projection; the dashed lines are latitude circles, the solid lines delineate the region over which the vorticity will be computed.

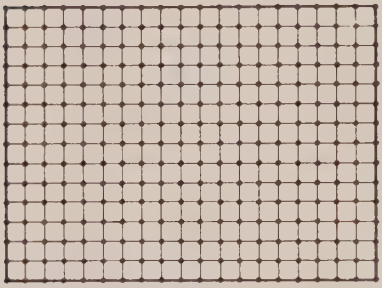


Figure 4. A network of grid points covers the rectangular region of figure 3. The grid shown here has 18 interior columns and 13 interior rows.

suited — namely those that involve extraordinary amounts of calculation.

Suppose we have an analyzed 500-mb map showing height contours. *Problem: to calculate the geostrophic vorticity field that corresponds to the contour analysis.* Let z denote contour height of the 500-mb surface, g the acceleration of gravity, and f the Coriolis parameter, so that the geostrophic wind components are

$$u = -\frac{g}{f}\frac{\partial z}{\partial y}; \qquad v = +\frac{g}{f}\frac{\partial z}{\partial x}.$$

Then the geostrophic vorticity is

$$\zeta = \frac{\partial v}{\partial x} - \frac{\partial u}{\partial y} = \frac{g}{f}\left(\frac{\partial^2 z}{\partial x^2} + \frac{\partial^2 z}{\partial y^2}\right), \qquad (1)$$

neglecting here the variation of f in the denominator of the expression for u.

The contour analysis is drawn on a map that represents a plane projection of the spherical earth; this projection will distort true distances. Let Δx be the true distance (on the earth) between two neighboring points and let Δa be the distance between these same points measured on the projection; then $\Delta a = m\,\Delta x$, where m is the *scale factor* for the projection and will usually depend upon the latitude where the measurement is made. To be specific, assume that a polar stereographic projection is used; the scale factor for this projection is

$$m = \frac{2}{1 + \sin\Phi},$$

where Φ is the latitude. Writing $\Delta a = m\,\Delta x$ and $\Delta b = m\,\Delta y$ for distances measured on the projection, the equation for calculating the vorticity from a contour analysis drawn on the projection is, from (1)

$$\zeta = \frac{m^2 g}{f}\left(\frac{\partial^2 z}{\partial a^2} + \frac{\partial^2 z}{\partial b^2}\right). \qquad (2)$$

It will be recalled that the Coriolis parameter f is equal to $2\omega \sin\Phi$, where ω is the angular speed of the earth's rotation.

Figure 3 is a portion of a polar stereographic projection on which a rectangular boundary has been marked out, to delineate the region over which the geostrophic vorticity will be computed. (The contour analysis is not shown.) We now construct a network of grid points (figure 4) covering the rectangular region, such that the distance between grid points is the same in either diretion — say equal to d. In order to specify the location of a grid point in this array, let i designate the *column* in which the point is located (counting from left to right), and j the *row* (counting from bottom to top); thus, $i = 0$ designates the column coinciding with the left-hand boundary, $i = 1$ the first interior column, $i = 2$ the next, and so on; similarly, $j = 0$ designates the row coinciding with the bottom boundary, $j = 1$ the row next above this, and so on.

From the contour analysis, values of the height z must be interpolated (this can be done visually) for each of the grid points. The interpolated height at the grid point whose coordinates are i, j we denote by z_{ij}. Consider a typical group of five neighboring

grid points (figure 5), the central point with coordinates i, j; we wish to calculate the geostrophic vorticity ζ_{ij} at this central point. Returning to formula (2) we see that the vorticity depends upon g (a constant, the same for all points), upon the scale factor m_{ij} at this point, the Coriolis parameter f_{ij} at this point, and the second derivatives of the contour height in both directions at this point.

To evaluate the latter two terms, consider the point half way between the central grid point (i, j) and the next grid point to the right $(i+1, j)$; there, the *first* derivative $\partial z/\partial a$ is approximately

$$(z_{i+1,j} - z_{i,j})\ d,$$

where d, as previously stated, is the distance between neighboring grid points. Likewise, half way between i, j and $i-1, j$ the first derivative $\partial z/\partial a$ is approximately

$$(z_{i,j} - z_{i-1,j})/d.$$

The *second* derivative $\partial^2 z/\partial a^2$ at the central point (i, j) is then approximately the difference between these two first derivatives, divided by d (the distance between the two "half-way" points); this reduces to

$$(z_{i+1,j} + z_{i-1,j} - 2z_{i,j})\ d^2.$$

The second derivative $\partial^2 z/\partial b^2$ is approximated in the same way, by considering the points $i, j+1$ and $i, j-1$ above and below the central point; this gives

$$(z_{i,j+1} + z_{i,j-1} - 2z_{i,j})/d^2.$$

Adding the two second derivatives, as required in formula (2), we get

$$(z_{i+1,j} + z_{i,j+1} + z_{i-1,j} + z_{i,j-1} - 4z_{i,j})/d^2.$$

The terms in parentheses consist simply of the sum of the height values at the four points surrounding the central point, minus four times the height at the central point; for brevity we will denote this group of terms by S_{ij}.

The vorticity at i, j is now to be calculated from the formula

$$\zeta_{ij} = \frac{m_{ij}^2 g}{f_{ij} d^2} S_{ij}, \qquad (3)$$

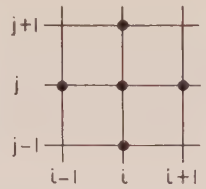


Figure 5. A typical grid point i, j and its four neighbors.

in which g and d are definite constants, m_{ij} and f_{ij} depend upon the latitude of grid point i, j (strictly, upon the sine of the latitude), and S_{ij} depends upon the contour heights at i, j and the surrounding four grid points.

There remains only to find the formulas by which m_{ij} and f_{ij} can be calculated for each i, j. Let r denote the linear distance between the pole and a point at latitude Φ on the polar stereographic projection; the properties of this projection require that

$$\sin \Phi = \frac{1 - (r/2a)^2}{1 + (r/2a)^2},$$

where a is the radius of the earth. If i_0, j_0 denote the grid coordinates of the pole, then the square of the linear distance (on the projection) between pole and grid point i, j is

$$r^2 = d^2\ [(i - i_0)^2 + (j - j_0)^2],$$

according to Pythagoras' theorem. The last two formulas provide for calculation of $\sin \Phi$ at any grid point i, j: we need only the values of i and j and the constants i_0, j_0, d, and a.

From the formulas already given, the reader can easily verify that the following computation scheme will yield the geostrophic vorticity ζ_{ij}:

Formulas:

$$S_{ij} = z_{i+1,j} + z_{i,j+1} + z_{i-1,j} + z_{i,j-1} - 4z_{i,j},$$

$$C_{ij} = (d/2a)^2\ [(i - i_0)^2 + (j - j_0)^2],$$

$$\zeta_{ij} = \frac{g}{2\omega d^2} \frac{(1 + C_{ij})^3}{1 - C_{ij}} S_{ij},$$

$$(i = 1, 2, 3, \ldots, p \text{ and } j = 1, 2, 3, \ldots, q)$$

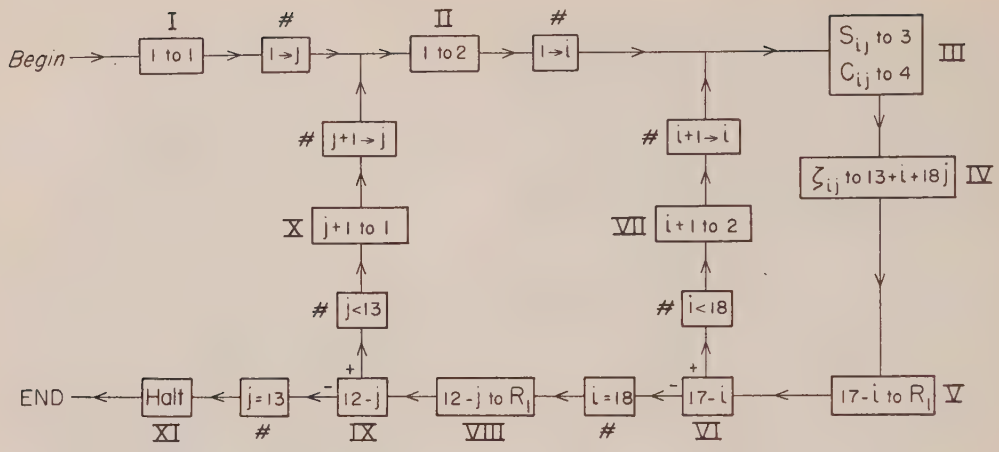


Figure 6. Flow diagram for vorticity computation. The formulas for S_{ij} and C_{ij} in box III and ζ_{ij} in box IV are listed in the program given previously.

Data:

z_{ij} = contour heights at all grid points.
($i = 0, 1, 2, \ldots, p+1$ and $j = 0, 1, 2, \ldots, q+1$)

Constants:

g = accceleration of gravity,
ω = angular speed of earth's rotation,
a = radius of earth,
d = distance between grid points on projection,
i_0, j_0 = grid coordinates of pole,
p = number of interior grid columns,
q = number of interior grid rows.

Roughly 10 multiplications are required to produce ζ_{ij} for one grid point from these formulae. Suppose, further, that something like a total of 100 orders (including nonarithmetic operations, such as necessary transfers of numbers between registers and memory) are required to produce one ζ_{ij}. Allowing 500 microseconds for each multiplication and 65 microseconds for each order, we get an estimate of $500 \times 10 + 65 \times 100 = 11{,}500$ microseconds, or say roughly 0.012 seconds as the time required to compute the vorticity at one grid point. For a grid of 500 points the whole computation would be finished in about $500 \times 0.012 = 6$ seconds — which justifies our earlier remark, that this particular problem is not representative of the type for which the high-speed computer should be used, where the computation time may be of the order of minutes or even hours.

3. A computation

The computation scheme outlined at the end of the preceding section constitutes what is known as the *program*. In order to portray the precise sequence of logical steps by which the control will carry out this program, we construct a *flow diagram*. The flow diagram for this computation is presented in figure 6.

Before interpreting figure 6, it is necessary to agree on a plan for enumerating the memory locations, so that the position of each number held in the memory can be specified. It will be recalled that the various digits belonging to one number are distributed among the memory tubes; each digit is assigned to one tube and the digits of one number appear in corresponding points of the charge networks of the various tubes (figure 1). Suppose each tube accommodates 1,024 points (charges) — so the capacity of the memory is 1,024 numbers — and that these points are arrayed in a square network of 32 rows and 32 columns. Each point in the array is a definite memory location. We enumerate these locations as follows: start in the lower left corner and call this point 0; proceed serially along the bottom row to 31 in the lower right corner; continue with the first (left-hand) point in the second row as 32 and proceed along this row to the end

point 63; likewise, 64 will be the first point in the third row, and so on; the very last point in the array (in the upper-right corner) will be number 1023, so that (since we started with 0) there are 1,024 locations in all.

Turning to figure 6, we note first that the flow diagram consists of a collection of "boxes" connected by branches marked with arrows. Each box (except those marked #) stands for a group of orders necessary to perform the operations indicated by notations within the box. The arrows on the branches connecting the boxes indicate the sequence in which the control will take up these groups of orders. Starting at the place marked "begin", the control proceeds to the first box, numbered I. The notation "1 to 1" in this box is intended to mean that the *number* 1 is ordered into *memory location* 1. The next box (marked #) does not stand for any operations or orders, but is intended to make an assertion as to how the number 1 introduced in the preceding box is henceforth to be interpreted; the notation "1 → j" signifies that we are to regard the number 1 (in memory location 1) as a value — in fact the first value — of the index j. This type of purely logical assertion is useful as an aid in reading and interpreting the flow diagram.

Before proceeding to box II, we remind the reader that i and j identify the column and row location of a point in the grid (not to be confused with the charge network in the memory!) that was marked out on the map projection (figure 4). Interpolated contour heights z_{ij} have been assigned to each of these grid points, and our objective is to compute the vorticity ζ_{ij} at each point. To fix ideas we will refer to a grid with 18 interior columns and 13 interior rows (figure 4), so that i runs from 0 (the left-hand boundary) to 19 (the right-hand boundary) and j runs from 0 (the bottom boundary) to 14 (the top boundary). This grid has $18 \times 13 = 234$ interior points; therefore, the results of the computation will consist of 234 values of ζ. The data required for the computation consist of the values of z at every grid point *including* points on the boundaries; therefore, $20 \times 15 = 300$ values of z comprise the data. (The four corner points are in fact not needed but it is convenient to include them in order that every row may have the same number of points.) In table 1 we have compiled a "storage table", which is a record of the memory locations assigned to the numerous quantities involved in the computation. The 234 values of ζ will be stored in locations 32 to 265, while the 300 values of z are stored in locations 288—587.

Table 1. Storage table.

Memory location	Quantity stored	Remarks (see fig. 6)
0	—	
1	j	Formed in boxes I and X
2	i	» » » II » VII
3	S_{ij}	Computed in box III
4	C_{ij}	» » » III
5—19	—	
20	0	
21	1	
22	$(d/2a)^2$	Used in box III
23	$g/2\omega d^2$	» » » IV
24	i_0	» » » III
25	j_0	» » » III
26	17	» » » V
27	12	» » » VIII
28—31	—	
32—265*	ζ	Computed in box IV
266—287	—	
288—587†	z	Data
588—607	—	
608— ?	Orders	
? —1023	—	

* ζ_{ij} is stored in location $13 + i + 18j$.
† z_{ij} » » » » $288 + i + 20j$.

Proceeding to box II, we have a group of orders that send the number 1 to memory location 2; the following box is an assertion that this number is to be interpreted as a value of index i. In box III we compute two numbers S_{ij} and C_{ij} needed in the formula for ζ_{ij} (note that i and j are the numbers in memory locations 2 and 1, respectively); S_{ij} is sent to position 3 and C_{ij} to position 4 in the memory. We can now calculate ζ_{ij} (box IV), which is sent to position $13 + i + 18j$; for example, $\zeta_{1,1}$ $(i = 1, j = 1)$ is sent to position $13 + 1 + 18 = 32$. The reader can easily verify that this scheme of storing the successively-computed values of ζ in the memory will place the required 234 numbers serially in memory positions 32 to 265 inclusive.[1]

[1] The last three paragraphs of this section give a more detailed account of the computations in boxes III and IV.

At this stage the computation of one of the ζ's is completed — namely, the one corresponding to $i=1$, $j=1$. The flow diagram is arranged in such a way that the ζ's are taken up in the following order: starting with $i=1$, $j=1$ we proceed along the first interior row to the last interior point ($i=18$, $j=1$), then to the first interior point of the next row ($i=1$, $j=2$) and along this row to $i=18$, $j=2$, and so on; the last ζ computed will be for $i=18$, $j=13$. Now the orders involved in boxes III and IV can be used for every ζ, so the control should return to box III each time we pass on to the next grid point. In order to control this iteration we must alter the values of i and/or j (in memory locations 2 and 1) so they will always correspond to the grid point whose ζ is being computed. Further, every time we complete the sequence of orders in boxes III and IV we must determine whether at that stage we are at the end of one of the interior rows ($i=18$), because if we are, one of two steps must be taken: if we are at the end of the *last* row ($j=13$), the iteration must be stopped because the computation is finished; if we are at the end of a row but *not* the last row, we must advance j and calculate for the next row, starting with $i=1$.

The orders that provide for this decision are indicated in boxes V and VI. In box V, the number $17 - i$ is sent to one of the registers, designated R_1. Box VI represents the *conditional transfer* order described previously. If $17 - i$ is positive (or zero) then $i < 18$ and we are not at the end of a row; in this case the control is transferred to box VII (along the "positive" branch of box VI), where the index i is advanced so that we proceed to the next point in the same row. From box VII the control is transferred *un*conditionally back to box III. When $17 - i$ first becomes negative we know that $i = 18$, the end of a row; we must then determine which of the two steps mentioned above must be taken, that is, we must determine whether we are on the last row. This is accomplished in box VIII, which orders $12 - j$ to register R_1, and box IX, which is the conditional transfer order. If $12 - j$ is positive (or zero) then $j < 13$ and we are not on the last row; in this case the control is transferred to box X (along the "positive" branch of box IX), where the index j is advanced, and then unconditionally to box II, where i is returned to its initial value 1, the first step in starting along the next row. When $12 - j$ first becomes negative we know that $j=13$, the last row, and that all ζ's have been computed. (Recall that box IX is reached through the "negative" branch of box VI, so it is entered only when $i=18$, the end of a row.) The "negative" branch ($j=13$) of box IX proceeds to box XI, a halt order.

It is evident that the flow diagram clearly portrays the sequence of operations by which the computation is performed, including the various points at which and to which the control must be transferred. The control cycles repeatedly around the loop (induction loop) formed by boxes III, IV, V, VI, VII, except that every 18th cycle carries it through the "larger" (less frequently traversed) loop formed by boxes III, IV, V, VI, VIII, IX, X, II.

The next step after completing the flow diagram is to write out the orders needed to perform the operations implicit in each box of the flow diagram. The complete list of orders required is the *code*. It will be recalled that the orders, properly translated into digits, will be stored in specific memory locations; referring to table 1, the reader will note that the orders for this computation are assigned to a portion of the memory beginning with location 608. The writer has not prepared the code corresponding to the flow diagram of figure 6, so the exact number of orders needed has not been determined.

There is one further aspect of the computation in boxes III and IV that should be mentioned. In box III the quantity S_{ij} is calculated, and this depends upon the values of z at i, j and at four points surrounding i, j (figure 5). The five values of z needed to compute S_{ij} must be called out of their memory locations by means of appropriate orders that will specify the "addresses" of these z's. Each time we pass through box III we are dealing with a different point (that is, a different pair of values i, j), which means that — as is obvious — a different set of five z's is required each time. Therefore, box III must include a provision for *changing addresses* in the orders that call out the z's; this is certainly feasible because each time we enter box III we know what the current values of i, j are (they are stored in locations 2 and 1), and we also know — according to the previously stated

scheme for storing the z's — that the address of z_{ij} is $288+i+20j$ (see table 1). This is an interesting and important feature of the flexibility of the control: by means of appropriate orders within the code, other orders can be changed whenever such changes are needed to suit the requirements of the computation.

In box IV a similar change of address must be included, in order to arrange for the storage of each newly-computed ζ_{ij} in its previously designated location. According to table 1, ζ_{ij} should be stored in location $13+i+18j$; therefore, before giving the order to store ζ_{ij}, the address in this order must be changed to correspond to the current values of i and j.

Finally, we may note that after S_{ij} and C_{ij} (computed in box III) are used to calculate ζ_{ij} in box IV, they will no longer be needed, so it is not necessary to store these two quantities permanently. In fact, each time S_{ij} and C_{ij} are computed and stored in locations 3 and 4, the values previously computed (and stored there) for the preceding point are automatically "erased".

4. Concluding remarks

The central problem of applied meteorology is weather forecasting in all its phases. One is naturally led to speculate on the extent to which high-speed automatic computers can be brought to bear on this problem. In the first place, it is necessary to repeat a comment made earlier: a computer of the type we are considering excels in dealing with computations that involve large quantities of data and (or), especially, prodigous amounts of arithmetic. The question is then, to what extent are advances in various phases of the problems of weather forecasting at present impeded by our inability to cope with difficulties of the type just mentioned? We will touch briefly only a few aspects of this question.

As a specific example, consider the problem of forecasting radiational temperature changes in the free atmosphere. A great deal is known about the theory of such changes: they are controlled by the distribution of water vapor, carbon dioxide, ozone, and other principal absorbers of long- and short-wave radiation, and by the distribution of liquid water (clouds), as well as the reflectivity and temperature of the underlying ground surface. Numerous graphical, tabular, and other approximate methods have been devised to account for most of these factors, but a more thorough treatment of the problem probably rests on detailed integration of the complete radiational transfer equations. The complexity of these equations is staggering, when one attempts to provide in detail for all of the factors that our present knowledge indicates are important, such as the distribution and intensity of lines in the spectra of the principal absorbers, and their variations with pressure and temperature, to mention only one element of the problem. A direct attack on the problem, attempting to account completely and simultaneously for all known factors, probably would be fruitless. However, it seems likely that a great deal could be learned by using a high-speed computer to integrate the transfer equations under conditions somewhat more complex than can reasonably be treated by means of "hand" computations.

As another example, take the forecast of high-level winds (over relatively short periods, say 24 hours or less) — for example, for use in operation of modern aircraft. Here we are face to face with the problem of predicting the future state of motion of the atmosphere — a problem inherent in virtually all phases of weather forecasting. From the time of Helmholtz, and particularly following the development of the Norwegian school of physical hydrodynamics under V. Bjerknes, meteorologists have on the whole had implicit faith in the belief that the laws of hydrodynamics, diligently applied to the atmosphere, will advance our understanding of the motion of the atmosphere and thereby improve our ability to predict its future state. Certainly it is true that our knowledge of the behavior of the atmosphere is vastly improved as a result of the approach through theoretical hydrodynamics, but because of the overwhelming complexity of the atmosphere, the advances as brought to bear on forecasting problems have been largely qualitative. It must be said that the meteorologist's faith in the efficacy of hydrodynamical methods has not been genuinely rewarded: modern forecasting techniques in most of the world's weather services probably are preponderantly subjective, and for the

most part do not reflect appreciably in quantitative terms, the advances in theoretical "dynamic" meteorology.

In these circumstances we are in fact encountering the difficulties mentioned at the outset of this discussion: strict application of hydrodynamical methods entails an unearthly amount of computation, as well as vast quantities of data, and is far beyond the reach of "hand" computation. Perhaps it is not unreasonable to advocate renewed faith in hydrodynamical methods of forecasting the future state of the atmosphere, because the high-speed computer may provide a means — hitherto unavailable — for helping to develop such methods into an effective quantitative implement.

At present the theoretical foundation of modern "dynamic" meteorology consists largely of classical hydrodynamics, amended let us say, as by Reynolds, in particular to include eddy transport phenomena. We cannot be certain that this framework will suffice ultimately to account for all of the complex phenomena observed in the atmosphere. To some extent it is reasonable to hold the view that modifications or reinterpretations of the classical framework will be suggested in the course of systematic scrutiny of the observations now or eventually at our disposal. However, no matter how thorough or inspired our interpretation of the observations may be, we shall not completely understand the limitations of the classical framework unless we learn a great deal about precisely what can be predicted within this framework; this of course means integration of the (nonlinear) hydrodynamical equations and use of high-speed computers.

In this connection, we should mention the recent and very promising revival of experimental methods of approaching atmospheric circulation problems by means of hydrodynamical models studied in the laboratory (notably the work of Fultz and his coworkers at the University of Chicago). The high-speed computer may prove effective in studying these controlled experiments theoretically, and if so, this coordinated approach may establish a more rigorous basis for testing the theoretical framework of our ideas about the atmosphere.

Acknowledgments. — On several occasions the writer has participated in the work of the Meteorology Group in the Electronic Computer Project at the Institute for Advanced Study, Princeton, and for these opportunities is sincerely indebted to Prof. J. von Neumann and Dr. H. H. Goldstine, and particularly also to Dr. Jule G. Charney. The writer acknowledges gratefully their advice concerning various aspects of this paper.

REFERENCES

There is a very extensive literature on the subject of automatic computers; but since the discussion in this paper pertains specifically to the logical aspects of the electronic digital computer at the Institute for Advanced Study, we cite only the following reports:

Burks, A. W., Goldstine, H. H., von Neumann, J., 1946: *Preliminary discussion of the logical design of an electronic computing instrument*, Princeton, Institute for Advanced Study.

Goldstine, H. H., von Neumann, J., 1947: *Planning and coding of problems for an electronic computing instrument*, Princeton, Institute for Advanced Study.

For an example of the formulation and programming of a concrete meteorological forecast problem, as well as a report on the results of computation with a high-speed computer, the reader should consult

Charney, J. G., Fjørtoft, R., von Neumann, J., 1950: Numerical integration of the barotropic vorticity equation. *Tellus*, **2**, 237—254.

On a Numerical Method of Integrating the Barotropic Vorticity Equation

By RAGNAR FJØRTOFT
University of Copenhagen

(Manuscript received 11 August 1952)

Abstract

A method is given by means of which quasi-permanent "velocity" fields may be constructed in which absolute vorticity is conserved if the atmosphere is treated as barotropic. For vorticity displacements these fields may be considered as constant for time intervals of the order of magnitude 24 hrs. Some simple approximate integration formulae are given for the Poisson- and Helmholz-equations. A graphical procedure is introduced to derive the isolines of vorticity and to perform the integration work.

1. Introduction

In numerical weather prognosis one works with a certain system of partial differential equations. These equations may differ considerably in form in accordance with how the problem has been formulated physically. Apart from this essentially *physical* aspect of the forecasting problem there is one *mathematical* and one *practical* aspect. As to the first one a number of mathematical problems arise in connection with an attempt of solving the equations. Certainly one is forced to solve the equations numerically, i. e. to replace in some way differentials with finite differences. How this has to be done must to some extent depend upon the physical assumptions which are made. However, even for a given system of equations different ways of numerical procedures may be tried, and it is not altogether certain which one should be preferred from a mathematical point of view. The partial differential equations in question are namely so complicated that a pure mathematical discussion of how well a numerical solution approximates the true one will have a limited value only. From a practical point of view, however, this may not present any great difficulty. The *relative* goodness of different numerical approaches may namely be tested from a comparison between the different solutions with respect to how well they approximate the actual development (integration) having taken place in the atmosphere.

The practical aspect of the numerical weather forecasting problem is related to the time necessary for carrying out a numerical solution. For it is quite clear that a forecast, how good it else may be, is completely useless in daily weather forecasting if it cannot be completed ahead of the weather. From that reason one may prefer one numerical method for another even on the cost of accuracy if the computation time is reduced essentially. This will especially be the case if the additional errors introduced thereby are well within the limits of errors present for other reasons.

In the following it is dealt with a particular numerical approach for solving the so-called barotropic vorticity equation. The chief achievements of this method may be summarized as follows:

1. For an area approximately as large as the one covered by the map in fig. 2 a forecast for 24 hrs. can be computed in the course of 2—3 hrs., reckoned from the conclusion of the analysis of the 500 mb map.

2. The work can be made entirely by one man and exclusively with the aid of standard means present at any weather forecasting section.

Recalling the integration time for the same problem on high-speed electronic computers, especially if the necessary introductory readings of data from the map are included, one would not expect particularly high integration accuracy from such a method. However, from a total of about 45 forecasts, including the four Princeton-forecasts (CHARNEY, FJØRTOFT, VON NEUMANN, 1950) it has turned out as a very pleasant fact that the integration accuracy appearingly is of the same order of magnitude as the one obtained on the electronic computer used for the Princeton-forecasts. It will be a subject for further work to make a systematic investigation of the integration accuracy. Such an investigation will be much easified after the completion of more forecasts based upon different numerical approaches. In spite of the fact that no systematic investigation of the integration accuracy has yet been carried through the author will dare to write down as a third achievement of the particular method in question:

3. In spite of the short integration time and in spite of the simple means by which the integration is carried through, the solutions obtained seem to be good approximations to the solutions of the barotropic vorticity equation.

In this article weight will be put on explaining the particular features of the method in question which are responsible for the surprisingly short integration time.

It is hardly worth while mentioning that it is the author's hope that weather forecasting centers will start using the method. This is not because of an opinion that forecasts based upon a barotropic model always will give useful results. Certainly it will be necessary later to include effects from the changes in the mass-field of the atmosphere. However, provided the numerical solutions may be considered as useful approximations to barotropic developments the meteorologists will themselves be able to encircle all the cases where barotropic forecasts fail and thereby gain valuable experience with regard to the importance of additional physical factors and how they possibly could be incorporated into the whole complex of computation and arguments building up a forecast. In this connection it should be mentioned that the numerical approach presented in this paper also will be valuable when applied to more extended physical models. This will be seen already in this paper from some arguments concerning the solution of the so-called Helmholz-equation $\nabla^2 \alpha - a^2 \alpha = h$ which appears in more extended models. Otherwise, this problem of using the particular numerical approach which is discussed in this article for less specialized models will be discussed more throughly in an article to appear later.

2. The barotropic problem

In the barotropic model it is assumed that the motion at levels close to the 500 mb level is horizontal and non-divergent thereby conserving the vertical component of absolute vorticity by the additional assumption of conservative forces. To reveal most clearly the essential features of the particular numerical method which is discussed in this article it will at first be assumed that the motion is a plane one. The fundamental property of periodicity around the earth for any quantity α derived from some atmospheric variable may in the plane representation be expressed by writing the quantity as a Fourier series. Since we are always dealing with numerical approaches to the solutions we will approximate the quantity in question by a finite Fourier series

$$\alpha = \sum^{(l_N)} \alpha_{m,n} e^{2\pi i \left(\frac{x}{\lambda_m} + \frac{y}{\omega_n}\right)} + c \tag{1}$$

where l_N now indicates the shortest wavelength occuring in the series and c is the constant in the series.

The law of individual conservation of vorticity may be written

$$\frac{\partial \hat{\eta}}{\partial t} = -\mathbf{v} \cdot \nabla \hat{\eta}, \tag{2}$$

$\hat{\eta}$ symbolizing vorticity, $\mathbf{v}$ velocity and ∇ the del-operator in the plane of motion. Because of the assumption of non-divergence for the velocity field, velocity may be expressed by a stream function ψ, $\mathbf{v} = -\nabla\psi \times \mathbf{k}$, $\mathbf{k}$ denoting a unit vector perpendicular to the plane of motion. Vorticity is then given by $\nabla^2\psi = \hat{\eta}$. The problem is now to solve (2) under the condition of periodicity expressed in (1). To get numerical solutions one will have to replace $\nabla^2\psi$ by an approximation, the corresponding finite difference expression. One may write

$$\hat{\eta} = \nabla^2\psi \approx -\frac{4}{d^2}(\psi - \overline{\psi}) \tag{3}$$

where $\overline{\psi}$ is defined from

$$\overline{\psi} = \frac{1}{4}\{\psi(x+d, y) + \psi(x-d, y) + \psi(x, y+d) + \psi(x, y-d)\}. \tag{4}$$

One may remark that even if $\nabla^2\psi$ is computed from the approximation (3) we do not necessarily require the solutions for gridpoints only. Even if a numerical solution cannot from evident reasons yield a solution everywhere we will to begin with in theory require a solution valid for any x and y. If now $\hat{\eta} = -\frac{4}{d^2}(\psi - \overline{\psi})$ is substituted into (2) and the factor $-\frac{4}{d^2}$ divided out of the equation one obtains as a new continuous problem to be solved under condition (1):

$$\frac{\partial \eta}{\partial t} = -\mathbf{v} \cdot \nabla \eta \tag{5}$$

where

$$\eta \equiv \psi - \overline{\psi}.$$

The numerical problem of solving (5) for a sufficiently short time $\triangle t$ may be divided into the two independent steps:

a) To displace properly the initial isolines $\eta = \text{const}$ over a time interval $\triangle t$ to find the new position of these lines and by subtraction the increment $\triangle\eta$ up to time $\triangle t$.

b) To solve thereafter $\triangle\psi - \overline{\triangle\psi} = \triangle\eta$ to find the corresponding increment in the stream-function of the velocity field.

These two steps will be discussed separately. At first, however, we shall in the next section discuss more closely the importance of the operation defined in (4) and some consequences which may be drawn from this operation.

3. Some consequences of the operation $\overline{}$

Performing the operation $\overline{}$ L times on each term of the Fourier representation (1) one obtains

$$\overline{\alpha}^{\}L} = \sum^{(l_N)} A_{m,n}^L \alpha_{m,n} e^{2\pi i\left(\frac{x}{\lambda_m} + \frac{y}{\omega_n}\right)} + c \tag{6}$$

with

$$A_{m,n} = \frac{1}{2}\left(\cos\frac{4d}{\lambda_m}\frac{\pi}{2} + \cos\frac{4d}{\omega_n}\frac{\pi}{2}\right). \tag{7}$$

From this last formula it may be concluded that for all components for which $4d/\lambda_m$ and $4d/\omega_n$ are not both equal to an even number one must have

$$|A_{m,n}| < 1.$$

The operation $\overline{}$ means therefore particularly a damping of all components with wavelengths $\lambda_m, \omega_n > 2d$. The damping is complete for $\lambda_m, \omega_n = 4d$, and is from here on decreasing asymptotically to no damping for $\lambda_m, \omega_n = 2d$, and to a reduced damping when λ_m, ω_n approach their highest values λ_1 and ω_1, respectively. By assuming $\lambda_m, \omega_n > 2d$ one obtains therefore by repeated $\overline{}$ operations in the limit:

$$\overline{\alpha}^{\}L} \underset{L\to\infty}{\longrightarrow} c, \quad \text{provided} \quad \lambda_m, \omega_n > 2d. \quad l_N > 2d. \tag{8}$$

The expression for $\alpha - \overline{\alpha}$ will in a finite Fourier series be given by

$$\alpha - \overline{\alpha} \equiv \sum^{(l_N)} B_{m,n} e^{2\pi i\left(\frac{x}{\lambda_m} + \frac{y}{\omega_n}\right)} \tag{9}$$

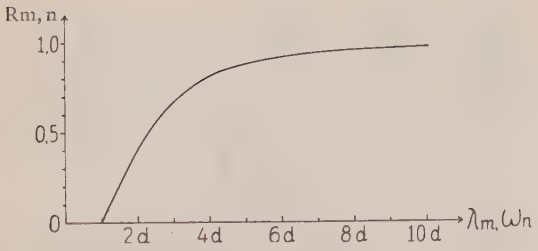


Fig. 1. The curve illustrates the goodness of the approximation (3) for different wavelengths.

where $B_{m,n}$ is given from

$$(10) \quad B_{m,n} = \left(\sin^2 \frac{4}{\lambda_m}\frac{d}{4}\frac{\pi}{4} + \sin^2 \frac{4}{\omega_n}\frac{d}{4}\frac{\pi}{4}\right) \alpha_{m,n}.$$

On the other hand $-\frac{d^2}{4}\nabla^2\alpha$ will as obtained from differentiation of (1) be given by

$$(11) \quad -\frac{d^2}{4}\nabla^2\alpha = d^2\pi^2 \sum^{(I_N)} \left(\frac{1}{\lambda_m^2} + \frac{1}{\omega_n^2}\right) \alpha_{m,n} e^{2\pi i\left(\frac{x}{\lambda m} + \frac{y}{\omega n}\right)}$$

The ratio $R_{m,n}$ between corresponding coefficients in (9) and (11) gives us then a measure for how well the approximation $\nabla^2\alpha = -\frac{4}{d^2}(\alpha - \overline{\alpha})$ is valid for the different components. We obtain $R_{m,n}$ from

$$(12) \quad R_{m,n} = \frac{\sin^2 \frac{4}{\lambda_m}\frac{d}{4}\frac{\pi}{4} + \sin^2 \frac{4}{\omega_n}\frac{d}{4}\frac{\pi}{4}}{d^2\pi^2/\lambda_m^2 + d^2\pi^2/\omega_n^2}$$

A curve is drawn in fig. 1 to illustrate the variation of $R_{m,n}$ for similar variations of λ_m and ω_n. As would have been expected $R_{m,n} < 1$, and is the closer to zero the smaller are the wave-lenghts. It is worth while mentioning, however, that it is the relative variation of $R_{m,n}$ and not the value of $R_{m,n}$ itself that introduces errors in the vorticity equation

$$\frac{\partial \hat{\eta}}{\partial t} = -\mathbf{v} \cdot \nabla \hat{\eta}$$

when $\hat{\eta}$ there is replaced by $-\frac{4}{d^2}(\psi - \overline{\psi})$ to give

$$\frac{\partial (\psi - \overline{\psi})}{\partial t} = -\mathbf{v} \cdot \nabla (\psi - \overline{\psi}).$$

The last equation may namely always be brought to give complete identity between the two equations for a single arbitrary component by multiplication with a suitable constant. The relative variation in $R_{m,n}$ is seen to be slight in the range of wavelengths $\lambda_m, \omega_n > 3\,d$, but some approximation to $\nabla^2\alpha$ is obtained even down to $\lambda_m, \omega_n = 2\,d$. In the following we shall fix the range of wavelengths in which we may expect useful solutions of (2) from the solutions of (5) to the range

$$\lambda_m, \omega_n > 3\,d.$$

4. The problem of time-integration

In Lagrangian terms the time-integration of $\frac{\partial \eta}{\partial t} = -\mathbf{v} \cdot \nabla \eta$ may in theory be accomplished by repeated displacements of the isolines in velocity fields which are considered as constant for a sufficiently small time increment $\triangle t$. This Lagrangian formulation of the time-integration problem will make sure that the fundamental assumption of conservation of vorticity is realized during the computation in the sence that no isolines η = const then will be present at any later time which did not already exist in the beginning, nor will any isolines be lost.[1]

The assumption of constant velocity fields during the partial displacements of the isolines of vorticity introduces necessarily in the general case an error in the distribution of η at a later time t. One may, however, keep this error within arbitrarily small limits by choosing a sufficiently small value for $\triangle t$. Decisive for how rapid a forecast can be computed for a period T is the ratio $\frac{\triangle t}{T}$. In the atmosphere the actual velocity fields are usually changing so rapidly that one is forced to keep $\triangle t$ less

[1] In this connection it should be recalled that in the Princeton-forecasts referred to earlier the vorticity was not always very well conserved in this respect.

than 3—4 hrs. It is possible, however, to increase this time considerably if use is made of the following considerations:
We write

$$\psi \equiv \psi^{\star} + \psi^{\star\star}$$

and correspondingly

$$\mathbf{v} = \mathbf{v}^{\star} + \mathbf{v}^{\star\star}$$

where

$$(13) \qquad \psi^{\star} = \psi - c\,(\psi - \overline{\psi}),\ c = \text{const}$$

$$\psi^{\star\star} = c\,(\psi - \overline{\psi}).$$

Substituting $\mathbf{v} = \mathbf{v}^{\star} + \mathbf{v}^{\star\star}$ into the vorticity equation

$$\frac{\partial\,(\psi - \overline{\psi})}{\partial t} = -\,\mathbf{v} \cdot \nabla\,(\psi - \overline{\psi})$$

one gets the equivalent equation

$$(14) \qquad \frac{\partial\,(\psi - \overline{\psi})}{\partial t} = -\,\mathbf{v}^{\star} \cdot \nabla\,(\psi - \overline{\psi}).$$

The term $-\,\mathbf{v}^{\star\star} \cdot \nabla \eta$ drops namely out of the equation because $\mathbf{v}^{\star\star}$ is parallel to the isolines $\eta \equiv \psi - \overline{\psi} = \text{const}$ and consequently ineffective in changing the distribution of vorticity. One may therefore as well displace the equiscalar lines of vorticity in the fictitious $\mathbf{v}^{\star}$-fields as in the actual velocity fields. This will now imply a considerable advantage if this fictitious velocity field $\mathbf{v}^{\star}$ is considerably less varying in time than the actual velocity field itself. It is quite clear that if the $\psi^{\star}$-field shall possibly possess this property, the constant c in (13) cannot be taken too small, nor can it be taken too large. If particularly c is put equal to 1 one obtains

$$(15) \qquad \psi^{\star} = \overline{\psi}.$$

Provided a useful displacement field exists it would imply a considerable advantage from a practical point of view if it could be derived from (15) instead of (13) because initially $\overline{\psi}$ already exists when $\eta_{t=0}$ is computed, and because at later time-steps the solution of $\triangle \psi - \overline{\triangle \psi} = \triangle \eta$ gives us immediately the change in the $\overline{\psi}$-field. The question then arises what size should be taken for the distance d involved in the definition of ψ to make $\overline{\psi}$ most useful as a displacement field. Obviously, since $\overline{\psi} \to \psi$ when $d \to 0$ only slight advantage could possibly be gained if d were taken too small. Recalling the consequences of the operation $\overline{}$ it is clear that in order to make $\overline{\psi}$ most suitable as a displacement field d should be taken equal to the order of magnitude of the wavelengths of the most rapidly changing components of the ψ-field. Applied to the atmosphere the value

$$4\,d \approx 2{,}500 \text{ km}$$

suggests itself as a natural value in this connection. In the forecasts referred to in the introduction d was put equal to the length of 12 degrees long. at 60 degrees N. It then turned out to be possible to take $\triangle t = 24$ hrs. Appearingly this only introduced negligible errors in most cases in the forecasted position of the equiscalar lines of vorticity. The map in fig. 2 illustrates a case of computed changes in $\triangle \eta \sim \triangle (z - \bar{z})$, where z is the height of the 500 mb surface. The isolines $\triangle (z - \bar{z}) = \text{constant}$ are drawn for every 40 m. The map in fig. 3 illustrates the corresponding observed changes. As it may be seen the observed and computed changes generally compare well. It should be remarked that the presented case is not one which is particularly favourable to the method of displacing described in this article because it was a case with relatively strong changes in the mean field itself. It should also be remarked that only relative vorticities were displaced, and also that discrepancies between observed and computed *vorticity* changes when they are of a disorderly character always have a tendency to disappear when the integration of $\triangle \psi - \overline{\triangle \psi} = \triangle \eta$ have transformed the vorticity changes into corresponding changes in the velocity field.

It should now be noted that having found a distance d involved in the definition of $\overline{\psi}$ which makes $\overline{\psi}$ most suitable as a displacement field, this distance may be too large for a sufficiently accurate determination of the vorticities from the approximation $\nabla^2 \psi = -\frac{4}{d^2}\,(\psi - \overline{\psi})$. Obviously if the distance d in this connection has to be taken much smaller, for instance equal to a $d_0 \ll d$, we have to return to the expression in (13) to get if possible

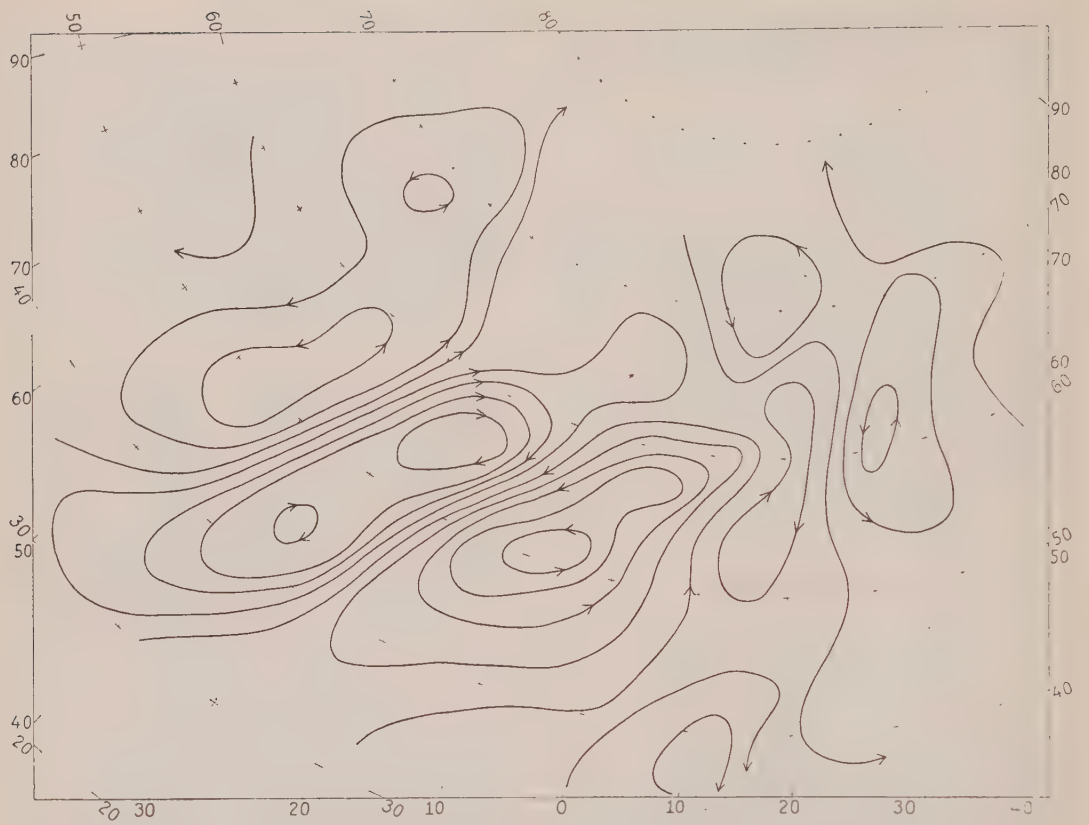


Fig. 2. November 25, 1951, 0300 GMT: The lines represent equiscalar lines for the computed 24-hr changes in $z - \bar{z}$, $\triangle (z - \bar{z})_{comp.}$, and are drawn for every 40 m. The arrows point in the directions $-\nabla \triangle (z - \bar{z})_{comp.} \times \mathbf{k}$.

a useful displacement field. We shall abstain from a discussion of what value should then at best be taken for the constant c in (13). Recalling namely what was pointed out at the end of the preceding section the approximation (3) will only be applicable for components with wavelengths down to approximately $3\,d$, which with the chosen value of $4\,d \approx 2{,}400$ km means down to wavelengths of the order of magnitude 1,800 km. Phenomena on a considerably smaller scale than this cannot easily be observed in the free atmosphere with any appreciable accuracy. It might also be that such phenomena, at least for relative short periods, do not have any noteable influence at all for the development of the more large-scale components of the motion. In that case there would also be a physical reason for abstaining from an attempt to include components with wavelengths considerably less than approximately 1,800 km.

5. The numerical solution of the Helmholz-equation $\nabla^2 \alpha - a^2 \alpha = -\hat{h}$.

When the changes $\triangle \eta$ in vorticity are found from the displacement of the isolines $\eta = \text{const}$, one is left with the problem of solving

$$\triangle \psi - \overline{\triangle \psi} = \triangle \eta .$$

This equation may be considered as a special case of

$$\alpha - M \bar{\alpha} = h \tag{16}$$

which, with M and h determined from

$$M = \frac{1}{1 + \dfrac{a^2 d^2}{4}} \tag{17}$$

and

$$h = -\frac{d^2 \hat{h}}{4 M},$$

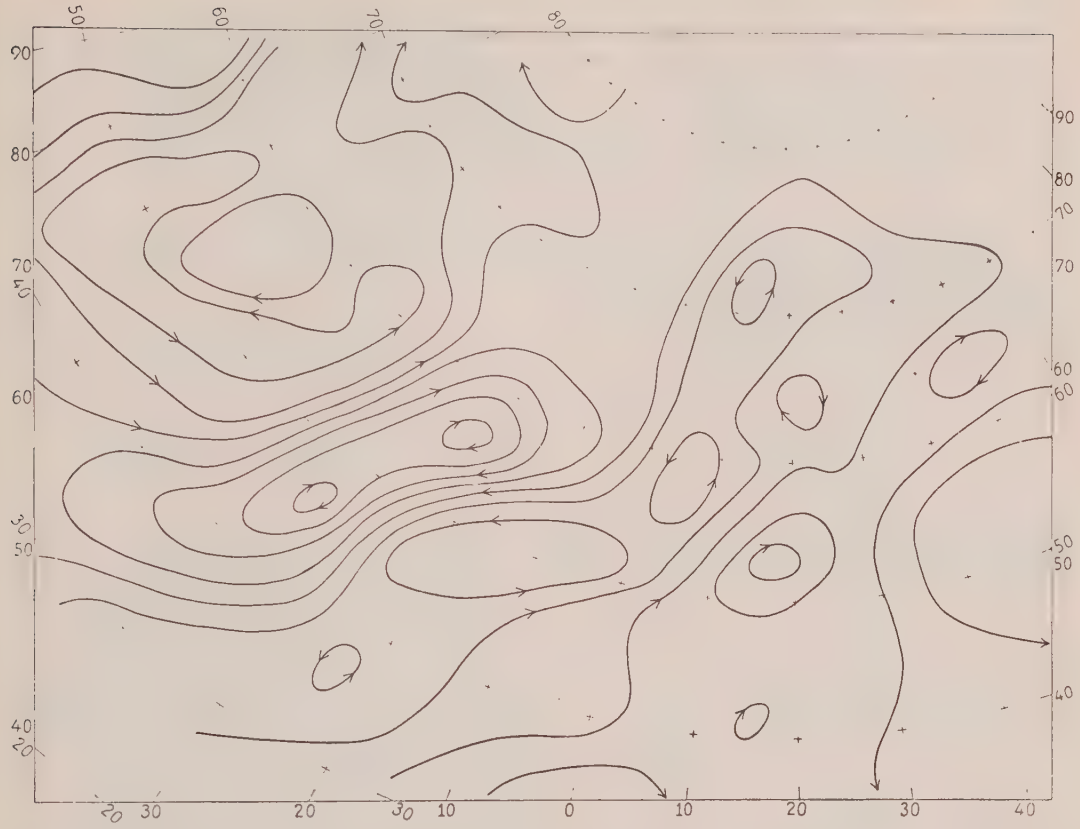


Fig. 3. November 25, 1951, 0300 GMT: The lines represent equiscalar lines for the observed 24-hr changes in $z - \bar{z}$, $\triangle (z - \bar{z})_{obs.}$, and are drawn for every 40 m. The arrows point in the directions $-\nabla \triangle (z - \bar{z})_{obs.} \times \mathbf{k}$.

is the finite difference equivalent to the Helmholz-equation

$$\nabla^2 \alpha - a^2 \alpha = -\hat{h}.$$

The following identity exists:

$$(18)\quad \alpha \equiv (\alpha - M\bar{\alpha}) + M(\bar{\alpha} - M\bar{\bar{\alpha}}) + \ldots + M^L\left(\bar{\bar{\alpha}}^{\}L} - M\bar{\bar{\alpha}}^{\}L+1}\right) + R_L$$

where

$$(19)\quad R_L = M^{L+1}\bar{\bar{\alpha}}^{\}L+1}$$

We are looking for a solution of (16) where the required quantity α is expressed as a finite Fourier series

$$\alpha = \sum^{(l_N)} \alpha_{m,n} e^{2\pi i\left(\frac{x}{\lambda_m} + \frac{y}{\omega_n}\right)} + c$$

with $\qquad l_N > 3\,d.$

The last restriction on l_N results from the fact mentioned at the end of section 1 that the approximation $\nabla^2 \alpha \approx -\frac{4}{d^2}(\alpha - \bar{\alpha})$ does not apply to the components with wavelengths considerably smaller than $3\,d$. We may now apply the result of (8) in section 1 in (19). We then obtain:

$$(20)\quad \underset{L\to\infty}{R_L \longrightarrow} 0, \qquad a^2 \neq 0; \quad M < 1.$$

and

$$(21)\quad \underset{L\to\infty}{R_L \longrightarrow} c, \qquad a^2 = 0; \quad M = 1.$$

We make further the assumption that the relative variation of $M = \frac{1}{1 + \frac{a^2 d^2}{4}}$ as compar-

ed with the one of $\overline{\overline{\alpha}}^{\,p}$ is sufficiently small to secure the approximate validity of

$$(22) \qquad \overline{M\overline{\alpha}}^{\,p} = M\overline{\alpha}^{\,p+1}$$

if a^2 is not already a constant. In the case of a Poisson equation, $a^2 = 0$, and (22) is authomatically satisfied. The identity (18) may therefore by applying (22) be written

$$(23) \quad \alpha \equiv (\alpha - M\overline{\alpha}) + M(\overline{\alpha - M\overline{\alpha}}) + \ldots + M^L \overline{(\alpha - M\overline{\alpha})}^{\,L} + R_L$$

However, $\alpha - M\overline{\alpha}$ is equal to the known quantity h in (16). We may therefore also write (23) as

$$(24) \quad \alpha = h + M\overline{h} + \ldots\ldots + M^L \overline{h}^{\,L} + R_L$$

Applying now (20), (21), the solution for α may be written as the infinite series

$$(25) \quad \alpha = h + M\overline{h} + \ldots + M^L \overline{h}^{\,L} + \ldots + c$$

where $c = 0$, when $a^2 \neq 0$

and $c =$ arbitrary, when $a^2 = 0$.

The arbitrary constant occurring in the case of a Poisson-equation is of course of no physical importance, as we, for instance, may understand when α is equal to the stream function. In the case of a Poisson-equation $\alpha - \overline{\alpha} = h$ it is immediately understood that the function h cannot contain any constant in the Fourier-development. This may, however, be the case when we have a Helmholz-equation $\alpha - M\overline{\alpha} = h$. The solution for α will in this case contain a constant determined from the solution (24). By the application of the Helmholz-equation in the more extended computational models of the atmosphere α is the temperature-change. From physical reasons, now, it may easily be understood why temperature in the mean may increase or decrease in a horizontal level extending all around earth.

The question about the rapidity of the convergence for each Fourier component of the infinite series (24) may be discussed easily by means of the expression (19) for the remainer R_L in (24).

Using (6) one obtains

$$(26) \quad R_L \equiv \sum^{(ln)} (R_L)_{m,n} e^{2\pi i \left(\frac{x}{\lambda_m} + \frac{y}{\omega_n}\right)} = \sum^{(l_n)} \alpha_{m,n} (A_{m,n} \cdot M)^{L+1} e^{2\pi i \left(\frac{x}{\lambda_m} + \frac{y}{\omega_n}\right)}$$

where now $A_{m,n}$ is given from

$$(7) \quad A_{m,n} = \frac{1}{2}\left(\cos \frac{4d}{\lambda_m}\frac{\pi}{2} + \cos \frac{4d}{\omega_n}\frac{\pi}{2}\right).$$

It is easily understood that the convergence is most rapid for the component with $\lambda_m, \omega_n = 4d$ in which case $A_{m,n} = 0$. From there on, however, the rapidity of the convergence is decreasing for λ_m, ω_n decreasing downwards to $3d$ which is the lower limit of the range of wavelengths we are interested in. The rapidity of the convergence is also decreasing steadily for λ_m, w_n increasing upwards from $4d$. As it may be concluded from (26) and (7) one must for instance take five terms in the infinite series (25) to get a solution in the Poisson case which is correct to 25 % in the whole range $4d < \lambda_m, \omega_n \leq 8d$. However, one can with fewer terms obtain solutions with approximate validity that approximate the true solutions more equally well over a wider range of wavelengths. The argument is as follows, taking for sake of illustration the Poisson case with $M = 1$:

Let now $\overline{\overline{\ }}^{\,q}$ denote any of the operations $\overline{\overline{\ }}^{\,1}, \overline{\overline{\ }}^{\,2}, \ldots, \overline{\overline{\ }}^{\,L}$ defined from

$$\overline{\beta}^{\,q} = \frac{1}{4}\{\beta(x + d_q, y) + \beta(x - d_q, y) + \beta(x, y + d_q) + \beta(x, y - d_q)\}$$

where

$$d_q < d_{q+1}, \quad q = 1, 2, \ldots, L.$$

Let us particularly define $\overline{\beta}^{\,0}$ from

$$\overline{\beta}^{\,0} = \beta.$$

We may then put down the following identity:

$$(27)\quad \alpha \equiv \sum_{q=0}^{L} \left\{ \begin{pmatrix} -q \\ \vdots \\ -1 \\ -0 \\ \alpha \end{pmatrix} - \overline{\begin{pmatrix} -q \\ \vdots \\ -1 \\ -0 \\ \alpha \end{pmatrix}}^{q+1} \right\} + r_L$$

where

$$(28)\quad r_L = \alpha^{\begin{matrix} -L+1 \\ -1 \\ -0 \end{matrix}}$$

Corresponding to the result (6), (7), in section 3 one must have

$$(29)\quad r_L = \sum^{(l_N)} \left(\prod_{q=1}^{L+1} A_{m,n}^{(q)} \right) \alpha_{m,n} e^{2\pi i \left(\frac{x}{\lambda_m} + \frac{y}{\omega_n} \right)}$$

where

$$(30)\quad A_{m,n}^{(q)} = \frac{1}{2} \left(\cos \frac{4 d_q}{\lambda_m} \frac{\pi}{2} + \cos \frac{4 d_q}{\omega_n} \frac{\pi}{2} \right).$$

For comparison the remainer R_L in (26) for the Poisson case with $M = 1$ is:

$$R_L = \sum^{(l_N)} (A_{m,n})^{L+1} \alpha_{m,n} e^{2\pi i \left(\frac{x}{\lambda_m} + \frac{y}{\omega_n} \right)}$$

While R_L is zero only for $\lambda_m, \omega_n = 4\,d$, r_L is seen to be zero for $\lambda_m, \omega_n = 4\,d_1, 4\,d_2, \ldots, 4\,d_{L+1}$. It would therefore so far be an advantage to use (27) instead of (18) as a base for obtaining approximate solutions of the Poisson equation $\alpha - \bar{\alpha} = 0$. It will then, however, further be necessary to express at least approximately the general term in the identity (27), after the remainder has been left out, in terms of the known function h. Denoting this general term with

$$P_q \equiv \begin{pmatrix} -q \\ \vdots \\ -1 \\ -0 \\ \alpha \end{pmatrix} - \overline{\begin{pmatrix} -q \\ \vdots \\ -1 \\ -0 \\ \alpha \end{pmatrix}}^{q+1}, \quad q = 0, 1, \ldots L$$

and letting further Q_q be defined from

$$Q_q \equiv \begin{pmatrix} -q \\ \vdots \\ -1 \\ -0 \\ \alpha \end{pmatrix} - \overline{\begin{pmatrix} -q \\ \vdots \\ -1 \\ -0 \\ \alpha \end{pmatrix}}^{q}, \quad q = 1, 2, \ldots, L$$

one obtains easily for the ratio between corresponding Fourier components of P_q and Q_q:

$$(31)\quad \frac{\{P_q\}_{m,n}}{\{Q_q\}_{m,n}} = \frac{1 - A_{m,n}^{(q+1)}}{1 - A_{m,n}^{(q)}} \equiv k_{m,n}.$$

Hence by substitution for $A_{m,n}^{(q)}$ and $A_{m,n}^{(q+1)}$ from (30), one obtains

$$k_{m,n} = \frac{\sin^2 \frac{4 d_{q+1}}{\lambda_m} \frac{\pi}{4} + \sin^2 \frac{4 d_{q+1}}{\omega_n} \frac{\pi}{4}}{\sin^2 \frac{4 d_q}{\lambda_m} \frac{\pi}{4} + \sin^2 \frac{4 d_q}{\omega_n} \frac{\pi}{4}}.$$

With the notation

$$\varkappa_q \equiv \frac{d_{q+1}}{d_q},$$

$k_{m,n}$ will for similar variations of λ_m, ω_n, say $\lambda_m, \omega_n = \lambda$, vary as

$$k_\lambda = \frac{\sin^2 \frac{\varkappa\, 4 d_q}{\lambda} \frac{\pi}{4}}{\sin^2 \frac{4 d_q}{\lambda} \frac{\pi}{4}}.$$

It is seen that

$$k_\lambda \underset{\lambda \to \infty}{\longrightarrow} \varkappa_q^2.$$

It is also understood that k_λ will be but slightly varying for large values of λ. If one wants k_λ as a quasi-constant for

$$\lambda_m, \omega_n > 4\,d_q,$$

one must take

$$\varkappa \leqq 1.5, \text{ approximately.}$$

In the following we shall take

$$(32)\quad \varkappa = 1.5.$$

We may then replace k_λ with the constant 2 which is approximately the mean value of $k_{4\,d_q}$ and k_∞:

$$k_{m,n} \approx 2, \quad \lambda_m, \omega_n \geqq 4\,d_q.$$

Consequently we may write (31) as:

$$(33)\quad P_q \approx 2\,Q_q \equiv P_q^\star \equiv \sum^{(l_N)} \{P_q^\star\}_{m,n} e^{2\pi i \left(\frac{x}{\lambda_m} + \frac{y}{\omega_n} \right)}$$

provided

$$\lambda_m, \omega_n \gtrless 4\,d_q. \tag{34}$$

We note from the definitions of Q_q and P_q that

$$Q_q = \overline{P}^{\,q}_{q-1} \tag{35}$$

Combining (33) and (35) one therefore gets:

$$P_q^{\star} = 2\,P_{q-1}, \quad \lambda_m, \omega_n \gtrless 4\,d_q.$$

By repeated use of this relation and condition (34) one finally obtains in connection with (33):

$$P_q^{\star} = 2^q \overline{P_0}^{\,q} \approx P_q \tag{36}$$

provided

$$\lambda_m, \omega_n \gtrless 4\,d_q.$$

Using

$$\overline{\{P_0\}}^{\,q}_{m,n} = \left(\prod_{r=1}^{q} A^{(r)}_{m,n}\right)\{P_0\}_{m,n}$$

in connection with (36), we may write:

$$\{P_q^{\star}\}_{m,n} = 2^q \left(\prod_{r=1}^{q} A^{(r)}_{m,n}\right)\{P_0\}_{m,n},$$

or, when similar variations of λ_m, ω_n are considered:

$$\{P_q^{\star}\}_{\lambda} = 2^q \cos\frac{4\,d_1}{\lambda}\frac{\pi}{2}\cdot\cos\frac{4\,d_2}{\lambda}\frac{\pi}{2}\cdot\ldots\cos\frac{4\,d_q}{\lambda}\frac{\pi}{2}\times \times \{P_0\}_{m,n}.$$

Consequently $P_q^{\star}$ is zero for $\lambda_m, \omega_n = 4\,d_1, 4\,d_2, \ldots, 4\,d_q$, and is also small for λ_m, ω_n having values between the values of $4\,d_1, 4\,d_2, \ldots, 4\,d_q$, if q is not taken to large. This is because the ratios $\frac{d_r}{d_{r-1}}$, $r = 2, \ldots, q$ have been chosen so small.

Making use of this last fact one may by substitution from (36) into the identity (27) write this approximately as:

$$\alpha \equiv P_0 + 2\,\overline{P_0}^{\,1} + \ldots + 2^L \overline{P_0}^{\,L-1} + r_L \tag{37}$$

for *all*

$$4\,d_1 \leq \lambda_m, \omega_n$$

if L is not taken too large. The relation (36) exists namely in parts of the range of wavelengths which are considered, and in the remaining parts where it does not exist the terms in consideration are negligible anyway in comparison with other more dominating terms. The remainder r_L given from (29), (30) will be small in the whole interval

$$4\,d_1 \leq \lambda_m, \omega_n \leq 4\,d_{L+1}.$$

By disregarding the remainder in (37) above and by substituting

$$P_0 = \alpha - \overline{\alpha}^{\,1} = h$$

one should therefore get an approximate solution α of the Poisson-equation $\alpha - \overline{\alpha}^{\,1} = h$ as:

$$\alpha^{\star} = h + 2\,\overline{h}^{\,1} + \ldots + 2^L \overline{h}^{\,L-1} \tag{38}$$

provided

$$4\,d_1 \leq \lambda_m, \omega_n \leq 4\,d_{L+1}.$$

The accuracy of the solution may be examined for the different Fourier components in the following way:

We have

$$h_{m,n} = \{\alpha - \overline{\alpha}^{\,1}\}_{m,n} = \left(1 - A^{(1)}_{m,n}\right)\alpha_{m,n}.$$

If this is substituted into (38) one obtains for each Fourier component:

$$\frac{\alpha^{\star}_{m,n}}{\alpha_{m,n}} = \left(1 - A^{(1)}_{m,n}\right)\cdot\left(1 + 2\,A^{(1)}_{m,n} + \ldots + 2^L A^{(1)}_{m,n}\ldots A^{(L)}_{m,n}\right). \tag{39}$$

When this ratio is $+1$ the solution is correct for the corresponding component. In the tabulation below this ratio is given as a function of $\lambda_m, \omega_n = \lambda$ measured in the unit $4\,d_1$, and for L assuming successively the values 1, 2, 3.

Table I.

	$4\,d_1$	$5\,d_1$	$6\,d_1$	$7\,d_1$	$8\,d_1$	$9\,d_1$	$10\,d_1$	$11\,d_1$	$12\,d_1$	$16\,d_1$
$\alpha^{\star}_{m,n}/\alpha_{m,n}$, $L=1$	1.00	1.12	1.00	.85	.82	.71	.62	.54	.36	—
$\alpha^{\star}_{m,n}/\alpha_{m,n}$, $L=2$	1.00	.82	.97	1.04	1.00	.92	.85	.77	.67	—
$a^{\star}_{m,n}/\alpha_{m,n}$, $L=3$	1.00	1.40	1.02	.85	.90	.89	.89	.88	.87	.76

It is seen that as approximate solutions of $\alpha - \overline{\alpha}^{\,I} = h$ we may use the following ones:

$$\alpha^{\star}_{I} = h + 2\,\overline{h}^{\,I}, \qquad 4\,d_1 \leq \lambda_m, \omega_n \leq 8\,d_1 \tag{40}$$

$$\alpha^{\star}_{II} = h + 2\,\overline{h}^{\,I} - 4\,\overline{h}^{\,I}, \qquad 4\,d_1 \leq \lambda_m, \omega_n \leq 10\,d_1 \tag{41}$$

$$\alpha^{\star}_{III} = h + 2\,\overline{h}^{\,I} + 4\,\overline{h}^{\,I} + 8\,\overline{h}^{\,I}, \qquad 4\,d_1 \leq \lambda_m, \omega_n \leq 16\,d_1. \tag{42}$$

If one wishes a solution with only two terms as in (40) which is better for wavelengths around $8\,d_1$ one could use a solution:

$$\alpha^{\star\star} = h + 3\,\overline{h}^{\,I}. \tag{43}$$

The goodness of this approximate solution is measured by

$$\frac{\alpha^{\star\star}_{m,n}}{\alpha_{m,n}} = \left(1 - A^{(1)}_{m,n}\right) \cdot \left(1 + 3\,A^{(1)}_{m,n}\right)$$

which is tabulated below:

Table II.

	$4\,d_1$	$5\,d_1$	$6\,d_1$	$7\,d_1$	$8\,d_1$	$9\,d_1$	$10\,d_1$
$\alpha^{\star\star}_{m,n}/\alpha_{m,n}$ ...	1	1.35	1.25	1.14	.94	.83	.69

To demonstrate the usefulness already of the approximation (43) a map for the 24 hr. *observed* changes $(\triangle z)_{obs}$ in the height of the 500 mb surface, fig. 4, was used to compute corresponding "Laplacians" $(\triangle z - \overline{\triangle z}) \equiv h$ which was then in turn substituted into formula (42) to give a set of integrated values $(\triangle z)^{\star\star}$, fig. 5. As is seen the correspondence is surprisingly good in spite of the fact that the case was not a particularly favourable one to the integration method.

It is worth while mentioning that in the general case of a Helmholz-equation $\alpha - M\overline{\alpha} = h$, the convergence for larger wavelengths may be speeded up for the series (24) by proceeding in a similar way as was done for the Poisson case. But in the case of a Helmholz equation this is not as necessary as in the case with a Poisson equation because the convergence for larger wavelengths have a tendency to be rapid already because of the multiplication with M^{L+1} in the remainer (26).

6. Modifications due to the curvature of the earth. — Geostrophic approximation

The curvature of the earth modifies the results of the preceding section physically and geometrically. Let now ∇_s denote the spherical del-operator and $\mathbf{k}$ a vertical unit vector. Relative velocity $\mathbf{v}$, and relative vorticity $\hat{\eta}$ will then be expressed by a stream function ψ as follows:

$$\mathbf{v} = -\nabla_s \psi \times \mathbf{k}$$

and

$$\hat{\eta} = \nabla_s^2 \psi.$$

If

$$f = 2\,\Omega \sin \varphi$$

where Ω is the angular velocity of the earth and φ is the latitude, the vertical component of absolute voticity $\hat{\eta}_{abs}$ will be given by:

$$\hat{\eta}_{abs} = \nabla_s^2 \psi + f.$$

The conservation of absolute vorticity may now be expressed by:

$$\frac{\partial \hat{\eta}_{abs}}{\partial t} = -\mathbf{v} \cdot \nabla_s \hat{\eta}_{abs}. \tag{44}$$

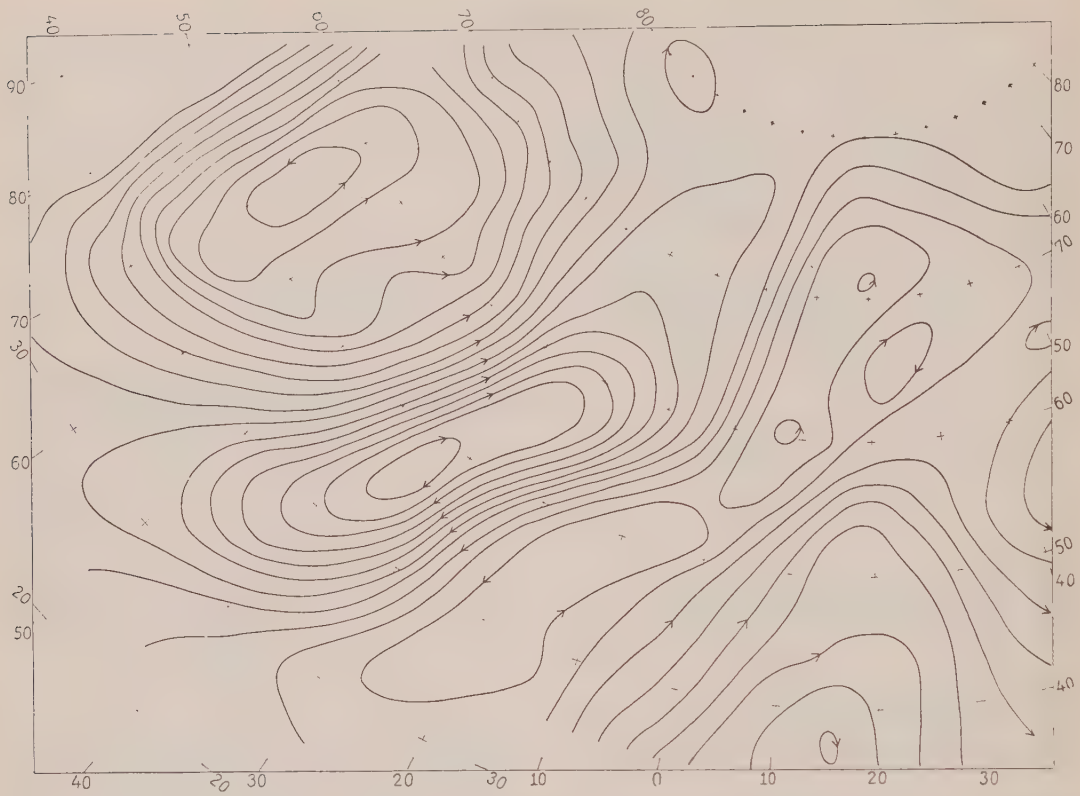


Fig. 4. November 25, 1951, 0300 GMT: The lines represent equiscalar lines for the observed 24-hr changes in z, $\triangle z_{obs.}$, and are drawn for every 40 m. The arrows point in the directions $-\nabla \triangle z_{obs.} \times \mathbf{k}$.

To find a useful displacement field one may proceed in exactly the same way as in section 4. We then write

$$\psi \equiv \psi^{*} + \psi^{**}$$

and correspondingly

$$\mathbf{v} = \mathbf{v}^{*} + \mathbf{v}^{**}$$

where

$$\psi^{*} = \psi - c\,(\nabla_s^2 \psi + f),\ c = \text{const}$$

and

$$\psi^{**} = c\,(\nabla_s^2 \psi + f).$$

Substituting $\mathbf{v} = \mathbf{v}^{*} + \mathbf{v}^{**}$ into (44) one obtains

$$\frac{\partial \hat{\eta}_{abs}}{\partial t} = -\mathbf{v}^{*} \cdot \nabla_s \hat{\eta}_{abs} \tag{45}$$

since the term $\mathbf{v}^{**} \cdot \nabla_s \hat{\eta}_{ab}$ drops out because $\mathbf{v}^{**}$ is parallel to the isolines of absolute vorticity and therefore ineffective in changing the vorticity distribution. Eq. (45) above therefore expresses the conservation of absolute vorticity in the field

$$\mathbf{v}^{*} = -\nabla_s \left\{\psi - c\,(\nabla_s^2 \psi + f)\right\} \times \mathbf{k}.$$

We now introduce the geostrophic approximation:

$$\mathbf{v} = -\frac{g \nabla_s z \times \mathbf{k}}{f}.$$

Hence relative vorticity will be given as

$$\nabla_s^2 \psi = \frac{g}{f} \nabla_s^2 z + \nabla_s \frac{1}{f} \times f \mathbf{v} \cdot \mathbf{k}$$

or approximately as

$$\nabla_s^2 \psi \approx \frac{g}{f} \nabla_s^2 z \tag{46}$$

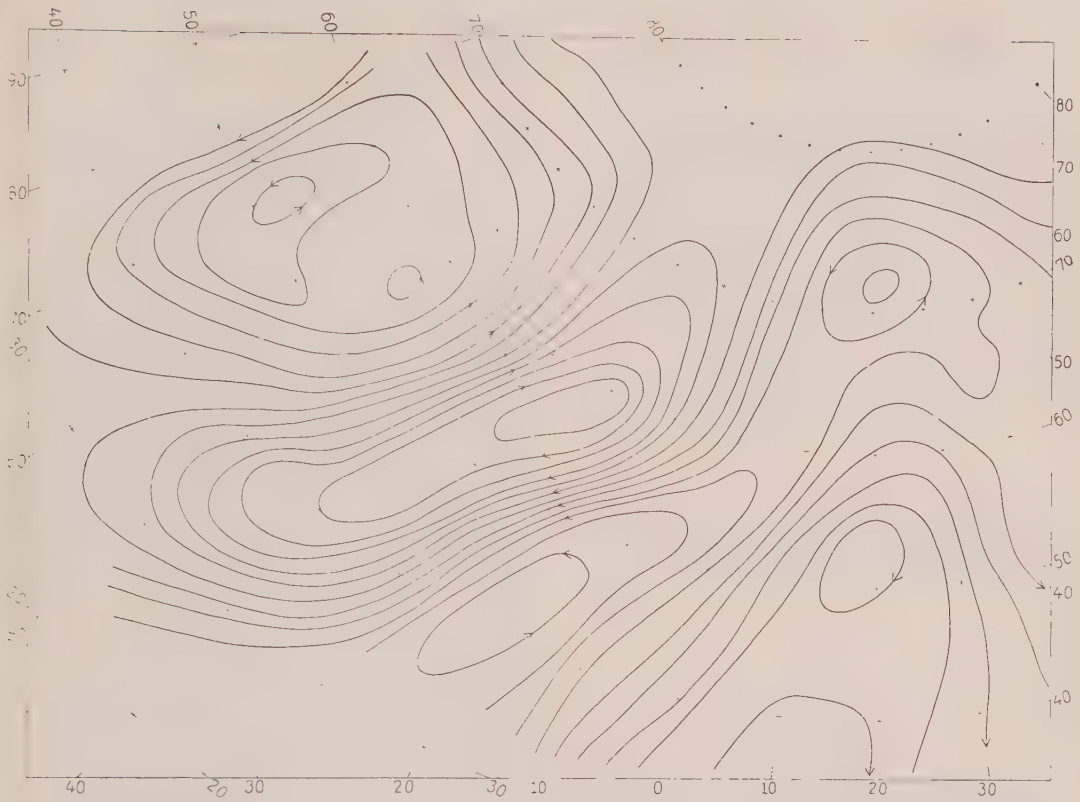


Fig. 5. November 25, 1951, 0300 GMT: The lines represent equiscalar lines for the 24-hr changes in z, $\triangle z_{int.}$, found from the approximate integration formula $\triangle z_{int.} = \triangle (z - \bar{z})_{obs.} + 3\ \overline{\triangle (z - \bar{z})}_{obs.}$. The arrows point in the directions $-\nabla\ \triangle z_{int.} \times \mathbf{k}$.

since $\frac{1}{f}$ may be considered as a virtual constant under the curl-operation $\nabla_s \times \mathbf{v} \cdot \mathbf{k}$. With (46) the vorticity-equation (44) can be written

$$(47) \qquad \frac{g}{f} \frac{\partial \nabla_s^2 z}{\partial t} = -\mathbf{v}_s \cdot \nabla_s \left(\frac{g}{f} \nabla_s^2 z + f \right).$$

Working in a conformal map projection with the magnification factor m, one obtains

$$\nabla_s^2 z = m^2 \nabla^2 z.$$

Here ∇ denotes the plane del-operator. Substituting the finite difference approximation

$$\nabla^2 z = -\frac{4}{d^2} (z - \bar{z}), \quad d = \text{const.}$$

in the last equation one obtains after division with the factor $-\frac{4g}{d^2}$ the finite difference equivalent of (47):

$$\frac{m^2}{f} \frac{\partial (z - \bar{z})}{\partial t} = -\mathbf{v} \cdot \nabla_s \left\{ \frac{m^2 (z - \bar{z})}{f} - \frac{d^2 f}{g\, 4} \right\}.$$

It may be supposed that the relative variation in $\frac{m^2}{f}$ is small compared with the relative variation in the quantity $z - \bar{z}$. The last equation may therefore also be written with sufficient accuracy:

$$(48) \qquad \frac{m^2}{f} \frac{\partial (z - \bar{z})}{\partial t} = -\frac{m^2}{f} \mathbf{v} \cdot \nabla_s (z - \bar{z}) + \\ + \frac{m^2}{f} \mathbf{v} \cdot \frac{f d^2}{m^2\, 4g} \nabla_s f.$$

Let us suppose that m is a function of latitude only. Then the last term in (48) may be written as

$$\frac{m^2}{f}\,\mathbf{v}\cdot\frac{4\Omega^2 d^2\,\sin\varphi\,\cos\varphi}{m^2(\varphi)\;4g}\,\nabla_s\varphi$$

or, with the notation $J(\varphi)$ defined from

$$(49)\quad J(\varphi)\equiv\int\frac{4\Omega^2 d^2\,\sin\varphi\,\cos\varphi\,d\varphi}{m^2(\varphi)\;4g}$$

as

$$\frac{m^2}{f}\,\mathbf{v}\cdot\nabla_s J(\varphi).$$

Substituting now this for the last term in (48) one obtains after dividing out the factor $\frac{m^2}{f}$:

$$(50)\quad \frac{\partial\xi}{\partial t}=-\mathbf{v}\cdot\nabla_s\xi,\qquad\text{where}$$

$$(51)\quad \xi\equiv z-\bar{z}-J(\varphi).$$

This equation, then, expresses the individual conservation of the quantity $\xi=z-\bar{z}-J(\varphi)$ in the geostrophic wind field $\mathbf{v}$.

To get an useful displacement field we now simply introduce in line with earlier considerations instead of $\mathbf{v}$ a fictitious velocity $\mathbf{v}^\star$ determined from:

$$\mathbf{v}^\star=-\frac{g}{f}\,\nabla_s z^\star\times\mathbf{k}$$

where

$$z^\star=z-c\left\{z-\bar{z}-J(\varphi)\right\},\quad c=\text{const.}$$

Choosing the constant equal to unity one gets:

$$z^\star=\bar{z}+J(\varphi)$$

and correspondingly

$$\mathbf{v}^\star=\bar{\mathbf{v}}+\mathbf{c}_d$$

with

$$(52)\quad \mathbf{c}_d=-\frac{g}{f}\,\nabla_s J(\varphi)\times\mathbf{k}.$$

Instead of (50) one obtains now the equivalent equation

$$(53)\quad \frac{\partial\xi}{\partial t}=-(\bar{\mathbf{v}}+\mathbf{c}_d)\cdot\nabla\xi.$$

This equation now expresses the conservation of the quantity ξ in the field $\bar{\mathbf{v}}+\mathbf{c}_d$. $\mathbf{c}_d$ as obtained from (52) and (49) is seen to represent a velocity directed westwards and is in fact the velocity of propagation of a so-called Rossby-wave with wavelengths approximately equal to $4\,d$. Since d has been taken to be only equal to the length of 12 degr. long. at 60 degr. N this wavespeed will be a small one and in most cases only imply a small correction to the field $\bar{\mathbf{v}}$. The main contribution from the variability in the Coriolis parameter will result from the term $+\bar{\mathbf{v}}\cdot\nabla_s J(\varphi)$ when the smoothed fields $\bar{\mathbf{v}}$ have noteable components in meridional direction.

7. Practical operation in a weather forecasting section

To accomplish a forecast in accordance with the principles outlined in this article it has from the following reasons turned out to be very convenient to make extended use of graphical additions and subtractions, rather than reading data in selected gridpoints only and furnish the necessary computational work with these data: *Firstly*, the analyzed map is in a sence utilized everywhere when graphical methods are used, whereas in a grid the gridpoints may have a very unfavourable position for a sufficiently accurate determination of the required quantities. *Secondly*, the graphical method is much faster than the grid method. For instance the whole time for the computation of a map for the isolines $z-\bar{z}=$ const reckoned from the conclusion of the analysis of the 500 mb map amounts to only 30—45 min, the area of the map being approximately as large as the ones covered by the maps presented in this article. *Thirdly*, one gets in the graphical method the isolines for the required quantities directly.

At a weather forecasting centre, the actual operational procedure then has to be as follows, or approximately as follows:

I: Construction of the initial field $\xi\equiv z-\bar{z}-J(\varphi)=$ const.

II: Displacement of the isolines $\xi = \text{const}$ in the displacement field $\overline{\mathbf{v}} + \mathbf{c}_d$, to find the new position of the lines after 24 hrs.

III: Solving the equation $\triangle z - \overline{\triangle z} = \triangle \xi$ to find the changes in the height z of the 500 mb map after 24 hrs.

I.

The 500 mb. map, here denoted as the z-map is copied over on a blank map displaced in an arbitrary direction x the distance d = equal to the length of 12 degr. long. at 60 degr. N. Only every second line is copied. This gives us a z_{x+d}-map for the isolines $z_{x+d} \equiv {}^1/_2 z(x+d, y) = \text{const}$. Thereafter is the z-map copied over once more on to the z_{x+d}-map, but now displaced a distance $-d$ in the x-direction. Again only every second isoline is copied. This gives us a map for the isolines $z_{x-d} = \text{const}$. By graphical addition of the z_{x+d}-map and the z_{x-d}-map one obtains when only every second line is drawn, a map representing the field

$$z^{\mathrm{I}} \equiv \frac{1}{4}\left\{z_{x+d} + z_{x-d}\right\} = \text{const.}$$

By repeating this for the direction perpendicular to the chosen x-direction one obtains similarly a representation for the isolines

$$z^{\mathrm{II}} = \frac{1}{4}\left\{z_{y+d} + z_{y-d}\right\} = \text{const.}$$

Then by the graphical addition of the z^{I}-map and the z^{II}-map above one obtains

$$z^{\mathrm{I}} + z^{\mathrm{II}} = \bar{z}.$$

Then finally from a graphical subtraction of the Z-field from the field $z - J(\varphi)$ one obtains the initial distribution of ξ given as isolines

$$\xi = z - \bar{z} - J(\varphi) = \text{const.}$$

It will obviously be most practical to have the isolines $J(\varphi) = \text{const}$ printed on blank maps once for all.

II.

The isolines $\xi = \text{const}$ are displaced in the field $\overline{\mathbf{v}}$ and then afterwards corrected for the relatively small westwards displacement due to the velocity $\mathbf{c}_d$. A wind-ruler of the ordinary hyperbolic type may be used. Instead of beginning with the displacement of the isolines $\xi = \text{const}$, one may at first draw a suitable line crossing all the isolines $z = \text{const}$ and then compute the displacement of this line for $L = \pm\triangle t,\ 2\triangle t, \ldots$, until the whole map is sufficiently densely covered of such lines. The displacement of the isolines $\xi = \text{const}$ will then afterwards be relatively easy.

III.

After having got the new position of the ξ-lines one obtains $\triangle\xi$ from the graphical subtraction

$$\xi_{24} - \xi_0 = \triangle\xi.$$

For integration of the equation

$$\triangle z - \overline{\triangle z} = \triangle\xi$$

one may use either of the approximate solutions (40), (43) or (41). Again, only graphical additions are involved in the computations.

8. Final remarks

In some future results obtained from the integration method in this article will be presented more systematically, also its extension to more extended models where absolute vorticity is not conserved. When the 45 forecasts referred to in the introduction were computed the integration formula (40) was used. It has later turned out that an use of the integration formula (41) instead, or even of (43), would have improved the integration results considerably in many cases. Also in most of the forecasts the influence from the variation of the Coriolis parameter was completely disregarded. It would therefore not be fair to the method to publish actual results at the present time. It will particularly be interesting to see how the forecasts compare with the ones obtained with electronic computers.

The time necessary for carrying out the three steps above amounts to between 2 and 3 hrs. After considerable experience one should expect that it will be possible to bring the time down to 2 hrs., particularly if one could use carbonpaper for the copying of the maps involved in the integration work. Having brought the time down to 2 hrs one could of course take the time and use systematically the more accurate integration formula (41). One could also possibly forecast for 48 hrs. if the barotropic model applies to such time intervals.

Acknowledgements. — The forecasts referred to in the introduction were prepared with the kind assistance from some of the staff at Videnskaps-Akademiets Institutt for Vær- og Klimaforskning, Oslo, and meteorologist Å. Moene at Meteorologisk Institutt, Oslo. I wish to express my best thanks to these people for their cooperation and good advices. I am also very much indebted to E. Høiland, the director of Videnskaps-Akademiets Institutt for Vær- og Klimaforskning, and to E. Grytøyr, the director of the weather forecasting section at Meteorologisk Institutt, Oslo, for their spontaneous willingness to allow for the practical arrangements in connection with a testing of the numerical method which has been discussed in this paper.

REFERENCE

J. G. Charney, R. Fjørtoft, J. von Neumann, 1950: Numerical integration of the barotropic vorticity equation. *Tellus* **2**, 4, pp. 237—254.

To the Comparison between Numerical Methods and Methods Now in Use for Forecasting Meteorological Charts

By OLOV LÖNNQVIST
Swedish Meteorological and Hydrological Institute

(Manuscript received 1 July 1952)

Abstract

Methods now in common use for making prognostic charts to a large extent are based on time derivatives, while the numerical methods start from space derivatives. That is the fundamental difference between them. It is shown that the present methods give at times very good results, and that the degree of success varies with the weather situation, characterized e. g. by the average interdiurnal pressure change over the test area. Because of this variation it would be desirable that every test of a numerical method includes a comparison with results obtained at some forecasting center or, at least, with the results of pure extrapolation.

Introduction

At present, in forecasting the sea-level pressure pattern as well as the upper flow pattern, in particular methods of extrapolation are used, i. e. the forecaster knows the previous history of the flow pattern, and he studies the changes or tendencies, the time derivatives. He may then introduce, in addition, certain corrections to the extrapolated pattern, argued from space derivatives, such as advection, divergence, vorticity etc., but extrapolation will always be the first step. It may be the areas of 24-hour changes which are moved either in accordance with their previous movement or along the track of the preceding changes, or it may be the troughs and ridges that are moved in a similar way. By extrapolation he can also make ridges build up or weaken and make troughs deepen or fill. Such a graphical extrapolation is rather easy to carry out, and it is not very subjective.

In general, however, extrapolation is not enough. The forecaster also must try to find signs of real evolutions, and then the methods grow more and more subjective. He has to look for areas of confluence or other kinds of advection of warm and cold air, and he has to estimate the effects to be expected from interactions between patterns over different parts of the map. He also has to apply rules of experience, general rules, rules valid for a certain season as well as special rules, just applicable to a prevailing type of weather.

But we should keep in mind that all these more subjective steps are just refinements, and that pure extrapolation already can give surprisingly good results.

In contrast to this the numerical methods, actualized by the advent of high-speed electronic computing machines, are all based on space derivatives, and the time derivatives

and the previous events do not enter at all. But at a meeting on numerical forecasting held in Stockholm in May 1952, there was briefly discussed if the time derivatives could not be entered in some way in the numerical prediction, e. g. to improve the analysis of the vorticity pattern in areas with sparse observations. Also the possibility of taking the previous events into account was mentioned, e. g. by introducing corrections in the forecasts for "past errors", as involving effects of neglected factors.

The development of the flow pattern is in present theories fully determined by the initial data, including all sorts of space derivatives. As however the time derivatives have proved invaluable for the non-automatic forecaster, it may be necessary to try to introduce them also in the numerical process. This ought to be done also if we do not want to replace the present methods entirely by numerical ones, but restrict ourselves to let the two kinds of methods supplement each other.

Present routines in forecasting

The goal for the meteorological prediction is in general to forecast the weather, i. e. surface wind, temperature, cloudiness and precipitation. It is true that forecast upper flow patterns may be used directly in preparing winds for aviation and ballistic purposes, but although this side of the problem is important, it will be neglected in the following.

To obtain a weather forecast we can either begin with an upper-air prediction and pass on to the sea-level and from there to the weather. Such a routine was recommended e. g. by RIEHL and collaborators (1951). Or we can begin with a sea-level prediction, as proposed by SCHERHAG (1948). In the latter case we can then pass to the weather almost directly, but our forecast will certainly improve if we construct an upper-air chart from the sea-level chart before making the weather forecast.

The advantage of the first way is that an aerological chart gives a much simpler picture of the weather situation than does the sea-level chart. The disadvantage is that upper-air charts must be translated into sea-level isobars and fronts before forecasting the weather.

The advantage of the sea-level way to the weather forecast, on the other hand, is partly the more direct relation between a sea-level chart and the weather, partly the existence of 3-hour tendencies on that chart. The disadvantage is the fact that sea-level prognostic charts can hardly be constructed directly for a longer time interval than 24 hours.

The choice between these two alternate ways depends on which one gives the best forecast in the actual case. Generally a compromise is the best way.

At the Aerological Division of the Swedish meteorological and hydrological institute the second method, in a form drawn up by SIMILÄ, is used for 24-hour forecasts, i.e. prognostic charts are prepared for sea-level pressure, relative topography and 500-mb topography in the said order. For 48-hour forecasts, on the other hand, the first method is used. Then the already prepared 24-hour 500-mb prognosis can serve as a hint.

Testing prognostic charts

Forecasts made by numerical methods have been tested by computing the coefficient for the correlation between computed and observed changes. Such tests have been performed by BOLIN & CHARNEY (1951), SAWYER & BUSHBY (1951) and STAFF MEMBERS, University of Stockholm (1952). It can be questioned if the correlation coefficient is the most appropriate number to characterize the degree of success in forecasting. Other numbers could be and have also been chosen for that purpose, such as the mean error, the ratio between mean error and mean change etc. However, we shall accept here the use of correlation coefficients, in spite of the fact that such a coefficient is not sufficient to state if a certain numerical forecast was successful. It is a common suggestion made by experienced forecasters that you must know in addition the corresponding coefficient for the forecast chart made for the same day by means of the old methods. This is necessary since the degree of difficulty in forecasting varies considerably from one situation to another.

When the Swedish routine is used, the correlation figure for the first prognosis, the 24-hour sea-level chart, will in general be higher than the figure for the secondary 24-

hour 500-mb prognosis. That will be shown in the following (see Table I). When the opposite routine is used, e. g. by numerical forecasting, we should then expect that the correlation figure for the 500-mb level will in general exceed the one for the sea-level chart with about the same amount. Therefore, if only 500-mb forecasts are compared by correlation methods, a considerably higher coefficient for the numerical forecast is required than for the routine one before we can say that the two methods will give just as good sea-level charts, which are in fact the most important ones for the forecaster.

As to the last step, the weather forecasting, we do not know exactly to what a degree the success in forecasting depends on the qualities of the prognostic chart. Hence it is not quite sure that improved prognostic charts, made in an automatic way, will improve automatically the weather forecasts. But to make the prognostic surface charts we need at present a good deal of reasoning, and it is quite sure that this reasoning is very important for the success in weather forecasting.

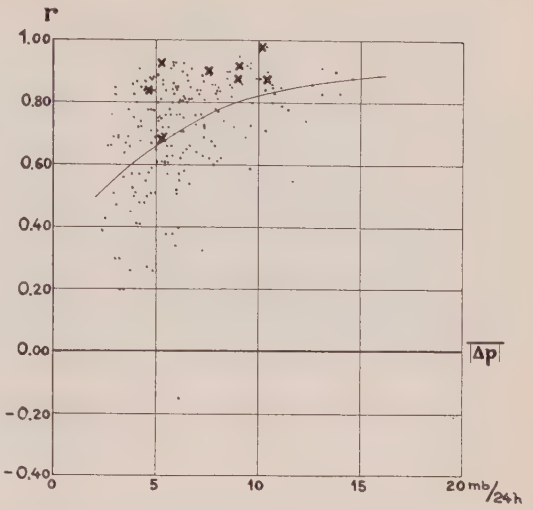


Fig. 1. Coefficients for the correlation between expected and observed pressure changes plotted as a function of the average diurnal pressure change over the test area. Crosses stand for cases studied in Table I. The curve represents average conditions.

Correlation coefficients obtained by present methods

For certain periods during the last two years coefficients have been calculated at the Swedish meteorological and hydrological institute for the correlation between expected and observed 24-hour changes in sea-level pressure. In all, 228 prognoses have been tested in that manner, and the obtained mean coefficient is 0.71. The success varied within wide limits. Thus the mean for the 30 cases in April 1951 was as high as 0.82. The variation may depend on varying ability of the forecasters. The variation due to the weather situation is probably more important. This effect has been studied in the following way (see fig. 1). In each one of the 228 cases, the coefficient of correlation, r, has been plotted against $\overline{|\Delta p|}$, i. e. the 24-hour change in sea-level pressure, averaged over the test area. This area covered northwestern Europe, and the number of points in the test grid was 26. The grid is shown in fig. 2.

We see from fig. 1 that forecasts of high quality are generally obtained when the average pressure change is large. When on the other hand the average pressure change is small, the success is highly variable, high correlations occur but also very low ones. In situations where the changes exceed in average 10 mb per 24 hours, the mean correlation is 0.85. The figures corresponding to the intervals 8—10, 6—8, 4—6 and 2—4 mb/24 hours are 0.83, 0.73, 0.69 and 0.55, respectively. The curve in fig. 1 represents this average relation between the correlation and the "intensity" of the weather situation, given in terms of the average diurnal pressure change over the test area.

The general experiences represented by fig. 1 are probably well-known, viz. it is easier to make a good prediction of the changes to occur when large changes are going on than in a flat and diffuse situation. But it is necessary to remind of the important fact that even if a prognostic chart gives a low correlation figure in a situation with small changes, the chart itself often looks fairly correct and is of good help in weather forecasting.

In 8 cases of the 228 ones, correlation coefficients were computed also for the changes in relative topography between 1,000 and 500 mb and for the changes in the 500-mb topography. These cases are denoted by crosses in fig. 1. The results are given in

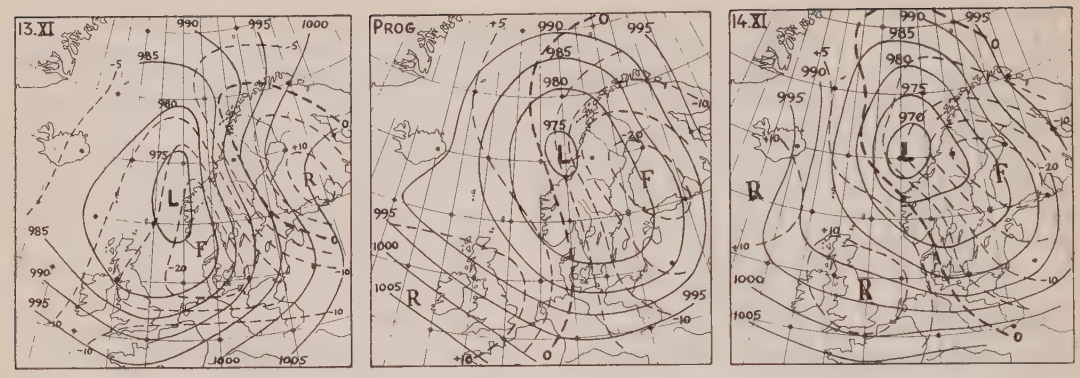


Fig. 2. Sea-level isobars (solid lines) and 24-hour isallobars (dashed lines) for the actual and prognostic chart of Nov. 13 and for the observed chart of Nov. 14, 1950. Contours labelled in mb. Correlation coefficient equal to 0.92. Test points inserted.

Table I. We see from the table that it seems to be easier to predict the sea-level pressure changes with accuracy than to predict the changes in the thickness pattern. We also see that there is in general, as mentioned above, a decrease in the agreement between expected and observed changes when the relative topography is added to the sea-level chart.

Table 1. Correlation between predicted and observed changes in sea-level pressure, relative topography and 500 mb topography.

Date	Sea-level	1,000—500 mb	500 mb
22—23 Oct. 1950	0.93	0.45	0.56
23—24 » 1950	0.84	0.76	0.84
24—25 » 1950	0.68	0.81	0.86
13—14 Nov. 1950	0.92	0.67	0.82
14—15 » 1950	0.88	0.18	**0.92**
10—11 Apr. 1951	0.90	0.73	0.75
11—12 » 1951	0.88	0.79	0.85
12—13 » 1951	**0.98**	**0.84**	0.78
Mean	0.88	0.65	0.80

Bold-face types are used for the best figure in each column. From the forecast for 12—13 April we see that the best sea-level forecast combined with the most successful relative topography forecast resulted in a 500-mb forecast which was one of the least successful of the forecasts studied here, and from the case of 14—15 November is seen that the best 500-mb forecast was obtained in spite of a failure in predicting the changes in the thickness pattern. This is peculiar, but we have not cases enough to see if it is just an accident. Anyhow, it demonstrates the complexity of the forecast problem but also, to some degree, the imperfection of the correlation method. In the case of 14—15 November the sea-level prognosis had its largest errors in some peripheric points in which the predicted changes in relative topography were also in error, while the agreement in the other points was good in the sea-level prognosis but moderate in the other one.

For the 48-hour changes in the 500-mb level the correlation coefficient has been computed in two cases, only, viz. 13—15 November and 14—16 November 1950. In both cases the figure was 0.79, i. e. a figure but slightly lower than the one for the 24-hour changes.

In Fig. 2, 3 and 4 is given an example of combined forecasts for sea-level pressure, relative topography and 500-mb topography. These forecasts were made on November 13, 1950, and the correlation coefficients are close to the mean values for the eight cases given in Table I. From fig. 2 we see that both the movement and the deformation of the principal fall-area are well predicted, but when the sea-level charts themselves are compared, we find a remarkable discrepancy as to the intensity of the low west of Norway, in spite of the high correlation 0.92. Regarding the relative topography we see from fig. 3 that the small changes were correctly sketched as to their main features. In fig. 4, giving the 500-mb charts, the movement of the pattern of 24-

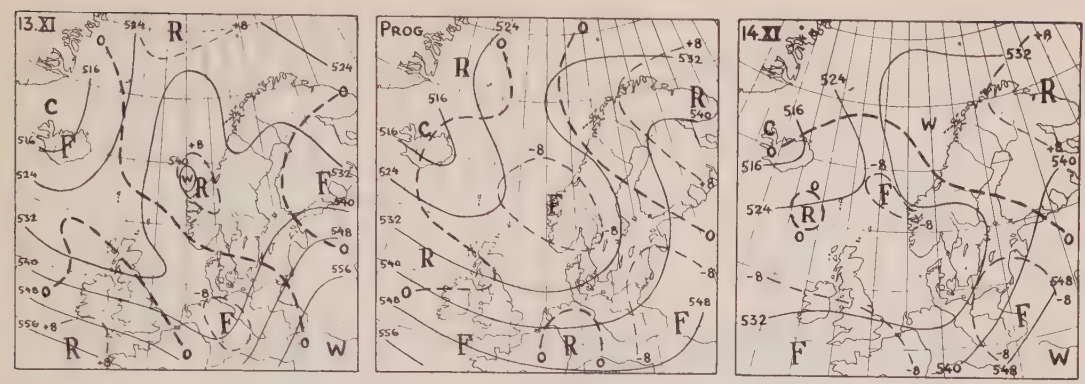


Fig. 3. 1,000—500 mb relative topography (solid lines) and 24-hour changes (dashed lines) for the actual and prognostic chart of Nov. 13 and for the observed chart of Nov. 14, 1950. Contours labelled in tens of metres. Correlation coefficient equal to 0.67.

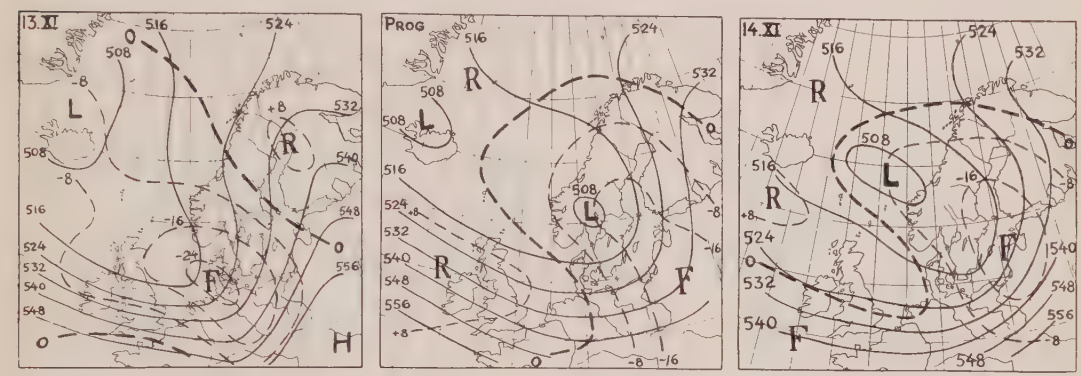


Fig. 4. 500 mb topography (solid lines) and 24-hour changes (dashed lines) for the actual and prognostic chart of Nov. 13 and for the observed chart of Nov. 14, 1950. Contours labelled in tens of metres. Correlation coefficient equal to 0.82.

hour changes turns out to be well estimated, giving an almost correct trough position. Over the southwestern part of the chart, however, there is a significiant discrepancy between the expected and observed upper flow pattern. Nevertheless the correlation coefficient is as high as 0.82.

Fig. 5 demonstrates an extraordinary good prognostic chart, tested in a somewhat different way. It is the 48-hour 500-mb forecast made December 17, 1951. It is tested in 54 points covering an area from Hudson Bay to Russia. The correlation coefficient was 0.92. If we leave out the nine westernmost points, the correlation goes up to 0.96. It is worth mentioning that in this case the situation was such that the forecast was made by pure extrapolation. Thus pure extrapolation can give a high correlation even for 48-hour changes over such a wide area, but the extremely good result is of course an exception. Nevertheless it demonstrates the well-known fact that the difficulty in forecasting meteorological charts varies considerably from one day to another. In this very case, numerical prediction must give just as high a correlation to be considered acceptable for practical use. In other cases, however, simple extrapolation gives perhaps not more than 0.20 correlation. Then the numerical prediction would be very successful even if the correlation is about 0.50. Therefore it is desirable that every test of a numerical forecast includes a comparison with result of pure extrapolation. Only such a comparison clearly points out the practical qualities of the numerical methods.

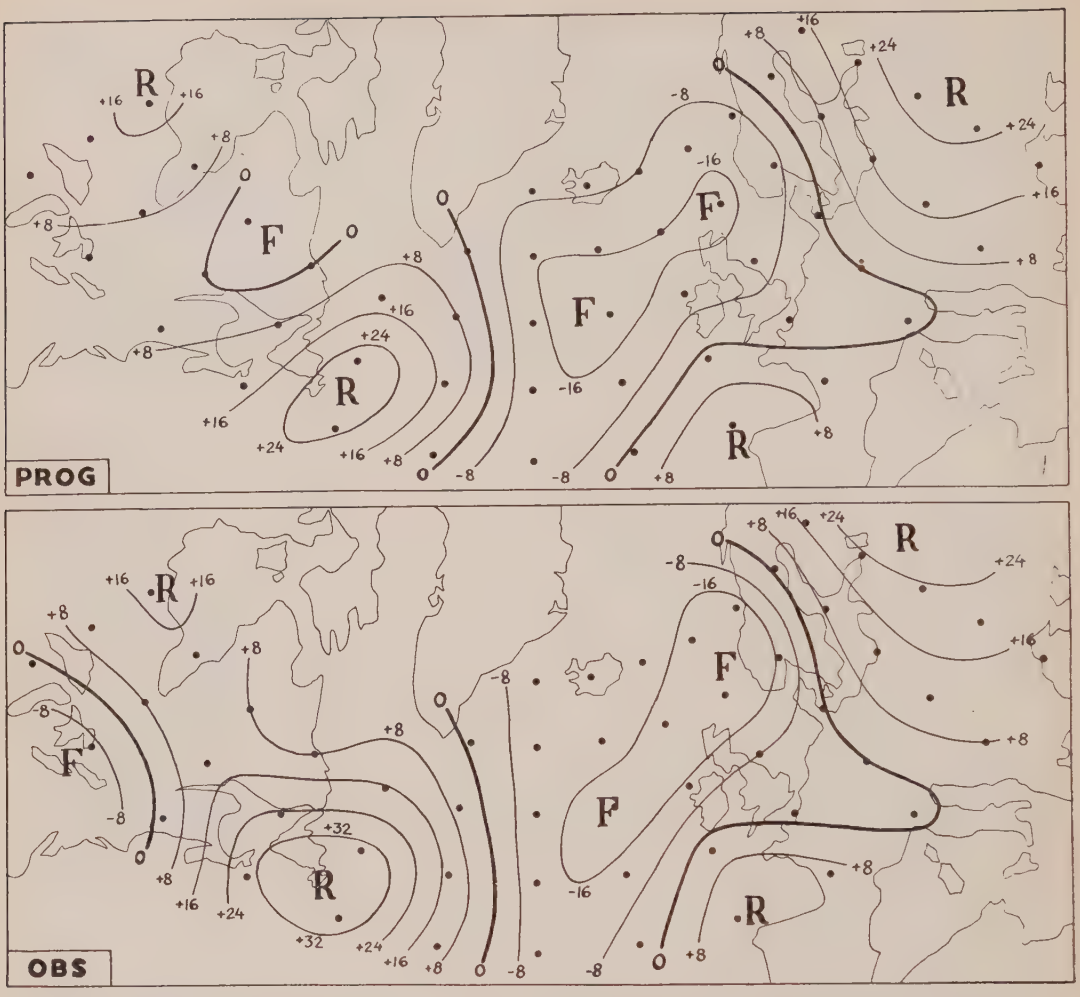


Fig. 5. A comparison between expected and observed 48-hour changes in 500-mb topography 17—19 Dec. 1951. Contours labelled in tens of metres. Correlation coefficient 0.92 computed from test points inserted.

REFERENCES

Bolin, B., and Charney, J., 1951: Numerical Tendency Computations from the Barotropic Vorticity Equation. *Tellus* **3**, 4, pp. 248—257.

Riehl, H., and, Collab., 1951: *Forecasting in Middle Latitudes.* Chicago. Univ. of Chic., Dep. of Met. 168 pp.

Sawyer, J. S., and Bushby, F. H., 1951: Note on the Numerical Integration of the Equations of Meteorological Dynamics. *Tellus* **3**, 3, pp. 201—203.

Scherhag, R., 1948: *Neue Methoden der Wetteranalyse und Wetterprognose.* Berlin—Göttingen—Heidelberg. Springer-Verlag. 424 pp.

Staff Members, University of Stockholm, 1952: Preliminary Report on the Prognostic Value of Barotropic Models in the Forecasting of 500 mb Height Changes. *Tellus* **4**, 1, pp. 21—30.

Gravitational Convection from a Boundary Source

By HUNTER ROUSE, C. S. YIH, and H. W. HUMPHREYS
Iowa Institute of Hydraulic Research, State University of Iowa, Iowa City

(Manuscript received 18 June, 1952)

Abstract

Elementary analyses of the mean patterns of free convection from a line source and a point source are presented without regard to the specific means by which the gravitational action is produced. The derived functional relationships are then verified and completed through use of velocity and temperature measurements above sources of heat, the generalized form of the results permitting characteristics of the mean flow to be determined over a considerable range of the primary variables. These results should enable meteorologists to evaluate the role of the basic convective process in the more complex movements of the atmosphere.

Introductory Remarks

Atmospheric disturbances are generally so complex in nature that the relative importance of the various factors which they involve can be appraised only by investigating the effect of each factor independently. In conducting an analysis of this nature, one often notes a close similarity between particular aspects of meterological and other flow occurrences, indicating that experience gained in related fields can be adapted to the problem in question. A case in point is the phenomenon of large-scale thermal updrafts in the atmosphere, one aspect of which is the comparatively simple process of gravitational or "free" convection from a boundary source, recently studied experimentally at the Iowa Institute of Hydraulic Research. The writers believe that the results of this investigation will permit meteorologists to evaluate the role of the basic convective mechanism in atmospheric phenomena.

Free convection due to a point source of heat is very simply illustrated by the plume of smoke which rises from a cigarette in otherwise stagnant air. Because of the buoyancy of the heated air in the immediate vicinity of the burning end, a continuous upward current is induced, with a corresponding radial inflow for reasons of continuity. While the initial steadiness of the smoke filament indicates purely laminar motion for a considerable distance above the heat source, the flow thereafter becomes unstable and the filament breaks up into the eddying clouds normally associated with turbulent motion. After the onset of turbulence, the effect of the convection upon the surrounding fluid becomes far more pronounced, the molecular shear and heat transfer of the initial laminar motion becoming dwarfed in scale by the macroscopic mixing action. Whereas the magnitude of the velocity along the vertical axis is then rapidly di-

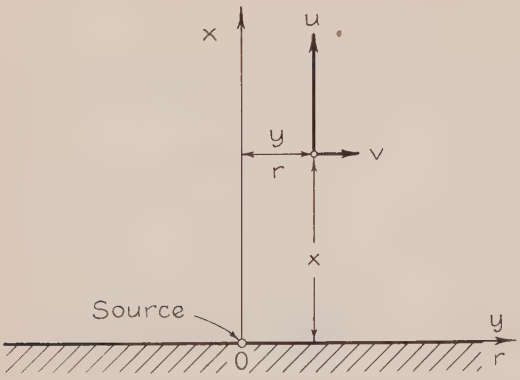


Fig. 1. Definition sketch.

minished, the angle of radial diffusion (and hence the rate of radial entrainment) is accordingly increased.

Although free convection of this nature is usually produced by a source of heat, it is the gravitational rather than the thermal aspects of the flow which are fundamental to the phenomenon. Indeed, essentially the same convection pattern can be produced in a liquid instead of a gas, and without any temperature change whatever. For example, the atmosphere may be replaced by a large body of water, and liquid having a different specific gravity may be introduced at a constant rate at a point on either horizontal boundary — a lighter liquid at the bottom or a heavier one at the top. In each event a similar pattern of convection will result, provided only that the local changes in density are relatively small in magnitude.

In the following pages an approximate analysis is made of the mean pattern of free convection from both a line source and a point source, without regard to the specific means of producing the gravitational action. Experimental evidence is then introduced to verify and complete the derived functional relationships, and generalized diagrams are presented to permit evaluation of the primary flow characteristics over a considerable range of the parameters involved.

Theoretical Analysis of the Mean Flow

If it is assumed that the scale of the initial laminar zone is small compared with that of the subsequent zone of turbulent convection, notation for either two-dimensional flow over a line source or axially symmetric flow over a point source may be reduced to that shown schematically in Fig. 1. Herein the origin O represents the location of either source. The local mean component of velocity in the vertical direction x is denoted by u, and that in the lateral direction y or radial direction r is denoted by v.

In addition to the mean velocity of convection above the source, there will also be a mean local change $\Delta\gamma$ in the weight density of the fluid. This entails, to be sure, a comparable change in the mass density ϱ. However, the following analysis is based upon the assumption that, whereas the unit buoyant force $-\Delta\gamma$ is sufficiently great to produce vertical acceleration, the corresponding variation in the mass density of the fluid undergoing acceleration is sufficiently small, in comparison with the density itself, to be neglected.

The foregoing assumption as to constancy of the density makes it possible to proceed in close accordance with the elementary analysis of flow in boundary layers, jets, and wakes. Three additional approximations are thus involved: the pressure intensity is assumed to be hydrostatically distributed throughout the field of motion; transverse forces are ignored in comparison with those in the vertical direction; and turbulent mixing in the vertical direction is ignored in comparison with that in the horizontal. The fundamental equations of motion thus reduce to the equation of continuity and simplified equations of vertical acceleration and diffusion. The development of the relationships describing convection above a line source will be described in some detail, but for purposes of brevity only the results will be indicated for the case of axial symmetry.

Under the foregoing assumptions, the equation of vertical acceleration in two-dimensional flow reduces to

$$\varrho u \frac{\partial u}{\partial x} + \varrho v \frac{\partial u}{\partial y} = -\Delta\gamma + \frac{\partial \tau}{\partial y} \qquad (1)$$

in which the intensity of vertical shear τ is proportional to the mean product of the turbulent velocity components u' and v':

$$\tau = -\varrho \overline{u'v'}$$

The equation of continuity is simply

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} = 0 \qquad (2)$$

Finally, the equation of diffusion of the quantity $\Delta\gamma$ has the form

$$u\frac{\partial(\Delta\gamma)}{\partial x} + v\frac{\partial(\Delta\gamma)}{\partial y} = -\frac{\partial w}{\partial y} \qquad (3)$$

in which the rate of lateral diffusion w due to turbulence is proportional to the mean product of the eddy scale l and the lateral velocity of fluctuation:

$$w = -\overline{v'l}\frac{\partial(\Delta\gamma)}{\partial y}$$

Since $u = 0$ at $y = \infty$, $v = 0$ at $y = 0$, and $\tau = 0$ at $y = 0$ and $y = \infty$, integration of Eq. (1), through use of Eq. (2), over a horizontal plane from the axis to infinity results in the following form of the momentum relationship:

$$\frac{d}{dx}\int_0^\infty \varrho u^2\, dy = -\int_0^\infty \Delta\gamma\, dy \qquad (4)$$

This states that the vertical gradient of the momentum flux is equal to the buoyancy of a horizontal stratum of unit thickness. As $\Delta\gamma = 0$ at $y = \infty$ and $w = 0$ at $y = 0$ and at $y = \infty$, a similar integration of Eq. (3) leads to the following expression for the constancy of the incremental weight flux past all successive planes:

$$\frac{d}{dx}\int_0^\infty u\,\Delta\gamma\, dy = 0 \qquad (5)$$

Finally, after multiplication of each term of Eq. (1) by u, integration of the resulting expression by parts, with the help of Eq. (1), yields the following form of the energy relationship:

$$\frac{d}{dx}\int_0^\infty \frac{\varrho u^3}{2}\, dy = -\int_0^\infty u\,\Delta\gamma\, dy - \int_0^\infty \tau\frac{\partial u}{\partial y}\, dy \qquad (6)$$

Because the quantity v^3 is negligible in comparison with u^3, this relationship states that the vertical gradient of the flux of kinetic energy is equal to the rate at which work is done by the buoyant force less the rate at which work is done by the turbulent shear. The last term at the same time represents the rate at which energy is lost to the mean motion through the generation of turbulence.

Although no indication is given as to the form of the functional relationships for u, $\Delta\gamma$, and τ which will satisfy these three necessary conditions, the customary hypothesis of dynamic similarity of the mean (and turbulent) motion at all elevations permits each of the relationships to be considered of the same form for every value of x. It is thus finally assumed that

$$\frac{u}{u_{max}} = f(\eta) \qquad \frac{\Delta\gamma}{\Delta\gamma_{max}} = g(\eta) \qquad \frac{\tau}{\varrho u_{max}^2/2} = h(\eta)$$

Herein $\eta = y/\sigma$, the quantity σ representing some linear characteristic of the velocity profile $u = u(y)$, such as the value of y at which the ratio u/u_{max} has some arbitrary magnitude.

Inasmuch as the unknown functions f, g, and h are independent of elevation, any power or combination thereof must also be constant with x. It is hence convenient to formulate a series of pertinent integrals for substitution into the three primary relationships which the functions must satisfy:

$$I_1 = \int_0^\infty f\, d\eta \qquad I_2 = \int_0^\infty f^2\, d\eta \qquad I_3 = \int_0^\infty f^3\, d\eta$$

$$I_4 = \int_0^\infty g\, d\eta \qquad I_5 = \int_0^\infty fg\, d\eta \qquad I_6 = \int_0^\infty h\frac{df}{d\eta}\, d\eta$$

Introduction of these terms into Eqs. (4), (5), and (6) permits the latter to be rewritten as

$$\frac{d}{dx}\left(I_2\,\varrho\, u_{max}^2\,\sigma\right) = -I_4\,\Delta\gamma_{max}\,\sigma$$

$$I_5\, u_{max}\,\Delta\gamma_{max}\,\sigma = C_x$$

$$\frac{d}{dx}\left(I_3\frac{\varrho\, u_{max}^3}{2}\sigma\right) = -C_x - I_6\frac{\varrho\, u_{max}^3}{2}$$

Simultaneous solution of these relationships will yield the following significant results:

$$\sigma = \frac{I_4 I_6 x}{2 I_2 I_5 - I_3 I_4} \sim x$$

$$u_{max} = \left(\frac{C_x}{\varrho} \frac{I_3 I_4 - 2 I_2 I_5}{I_2 I_5 I_6} \right)^{1/3} \sim x^0$$

$$\Delta\gamma_{max} = -\left(C_x \varrho^{1/2} \frac{I_3 I_4 - 2 I_2 I_5}{I_2 I_5 I_6} \right)^{2/3} \frac{I_2}{I_4 x} \sim x^{-1}$$

Evidently, no matter what forms the distribution curves may take, the convection zone will expand linearly with elevation; the maximum velocity (and hence the velocity along any line of constant y/x) will be independent of elevation; and the maximum incremental weight density (and hence that along any line of constant y/x) will vary inversely with elevation.

With this information it is possible to reach corresponding conclusions as to the variation of the volume flux Q, the momentum flux M, the kinetic-energy flux E, the unit buoyant force F, and the flux W of the incremental weight above a source of length L. Thus, per unit length of source,

$$\frac{Q}{L} = 2 \int_0^\infty u \, dy \sim x \qquad \frac{M}{L} = 2 \int_0^\infty \varrho u^2 \, dy \sim x$$

$$\frac{E}{L} = 2 \int_0^\infty \frac{\varrho u^3}{2} \, dy \sim x \qquad \frac{F}{L} = -2 \int_0^\infty \Delta\gamma \, dy \sim x^0$$

$$\frac{W}{L} = 2 \int_0^\infty u \, \Delta\gamma \, dy = 2 C_x$$

Herefrom it is seen that the flux of volume, the flux of momentum, and the flux of kinetic energy must increase continuously with elevation, for the force producing the convective motion is the same at all successive levels. As the flux of the incremental weight is also the same at every level, the quantity W must indicate the output of the source.

The corresponding analysis of the axially symmetric pattern of convection above a point source proceeds in a closely comparable manner. The resulting integral equations of momentum, diffusion, and energy have the forms

$$\frac{d}{dx} \int_0^\infty \varrho u^2 r \, dr = -\int_0^\infty \Delta\gamma \, r \, dr \tag{7}$$

$$\frac{d}{dx} \int_0^\infty u \, \Delta\gamma \, r \, dr = 0 \tag{8}$$

$$\frac{d}{dx} \int_0^\infty \frac{\varrho u^3}{2} r \, dr = -\int_0^\infty u \, \Delta\gamma \, r \, dr - \int_0^\infty \tau \frac{\partial u}{\partial r} r \, dr \tag{9}$$

Through the assumption of dynamic similarity of the mean flow at all elevations it is found that

$$\sigma \sim x \qquad u_{max} \sim x^{1/3} \qquad \Delta\gamma_{max} \sim x^{-5/3}$$

The following expressions are then obtained:

$$Q = 2\pi \int_0^\infty u \, r \, dr \sim x^{5/3} \qquad M = 2\pi \int_0^\infty \varrho u^2 r \, dr \sim x^{4/3}$$

$$E = 2\pi \int_0^\infty \frac{\varrho u^3}{2} r \, dr \sim x \qquad F = -2\pi \int_0^\infty \Delta\gamma r dr \sim x^{1/3}$$

$$W = 2\pi \int_0^\infty u \, \Delta\gamma \, r \, dr = 2\pi C_x'$$

Owing to the differences between these expressions and their counterparts for the two-dimensional case, basically different relationships are indicated for the flux of the several characteristic quantities past successive planes above a point source. Although the convection pattern again expands linearly with height above the source, the maximum velocity (and hence that for any constant value of r/x) now decreases continuously, and the incremental weight density decreases even more rapidly. On the other hand, the buoyant force on a stratum of unit thickness becomes steadily greater with increasing elevation, and Q, M, and E all increase at different rates. Once again, however, the flux of the incremental weight is the same at all elevations, so that W

must be equivalent to the output of the source.

Experimental Procedure and Results

In any experimental program to determine explicit relationships for either two-dimensional or axially symmetric convection from a boundary source, there are evidently four variables which can be controlled independently — two coordinate positions, the mass density of the fluid, and the strength of the source. These four quantities govern the magnitudes of such dependent variables as velocity, weight density, shear, turbulence, and their various combinations. While all of these factors would be of interest in a complete investigation of the phenomenon, attention is focused in the present study upon the major characteristics of the mean motion: the velocity component u and the change $\Delta\gamma$ in weight density. The resulting functional relationships between the dependent and independent variables for the two types of source can be indicated schematically as follows:

$$\frac{u}{\Delta\gamma} = f_{1,4}\left(x, \frac{\gamma}{r}, \varrho, \frac{W/L}{W}\right)$$

Inasmuch as five variables involving three fundamental dimensions appear in each of these four functions, it is to be expected that the dimensionless form of each will contain only two terms. Thus, for two-dimensional conditions, the relationships may be expressed as

$$\frac{u}{(W/L\varrho)^{1/3}} = \varphi_1\left(\frac{\gamma}{x}\right)$$

$$\frac{\Delta\gamma}{(\varrho W^2/L^2x^3)^{1/3}} = \varphi_2\left(\frac{\gamma}{x}\right)$$

For conditions of axial symmetry, the counterparts of these expressions are

$$\frac{u}{(W/\varrho x)^{1/3}} = \varphi_3\left(\frac{r}{x}\right)$$

$$\frac{\Delta\gamma}{(\varrho W^2/x^5)^{1/3}} = \varphi_4\left(\frac{r}{x}\right)$$

In each instance the grouping of terms is seen to be in accordance with the general relationships obtained analytically.

Experiments with heated air to determine the forms of these functions were conducted in closed rooms, particular care being required in the study of axially symmetric conditions to eliminate all sources of disturbance. In this instance the room was approximately circular in plan and had a diameter of 25 feet and an overall height of 11 feet. The heat source was placed at the midpoint of a centrally located platform 8 feet in diameter and a few inches in height. For the two-dimensional study the flow was confined between two parallel walls 4 feet high, 8 feet long, and 4 feet apart, and the source was placed across the midsection of a low platform extending the length of the walls. The heat sources in both cases consisted of recessed gas burners yielding low, blue flames approaching as nearly as practicable the desired point and line concentrations.

Measurements of the temperature distribution were made by means of a copper-constantan thermocouple, with its cold junction placed well away from the heat source, in combination with a potentiometer reading to 0.002 millivolt. Velocity indications were obtained with a specially constructed vane anemometer $1^1/_4$ inch in diameter, the jewel bearings of which permitted the indication of velocities as low as 0.2 foot per second. These instruments were mounted alternately on remotely controlled traversing mechanisms.

The rates of heat output from both the line and the point source were evaluated from the measured distributions of velocity and temperature according to the following thermal counterparts of the equations for W/L and W:

$$\frac{H}{L} = 2c_p\varrho\int_0^\infty u\,\Delta T\,dy \qquad (10)$$

$$H = 2\pi c_p\varrho\int_0^\infty u\,\Delta T\,r\,dr \qquad (11)$$

for the closely approximate condition that $\Delta\gamma/\gamma = -\Delta T/T$, the corresponding value of W in either case was computed from the resulting conversion equation,

$$W = -\frac{gH}{c_pT} \qquad (12)$$

Supplementary tests performed during the axially symmetric study were directed toward

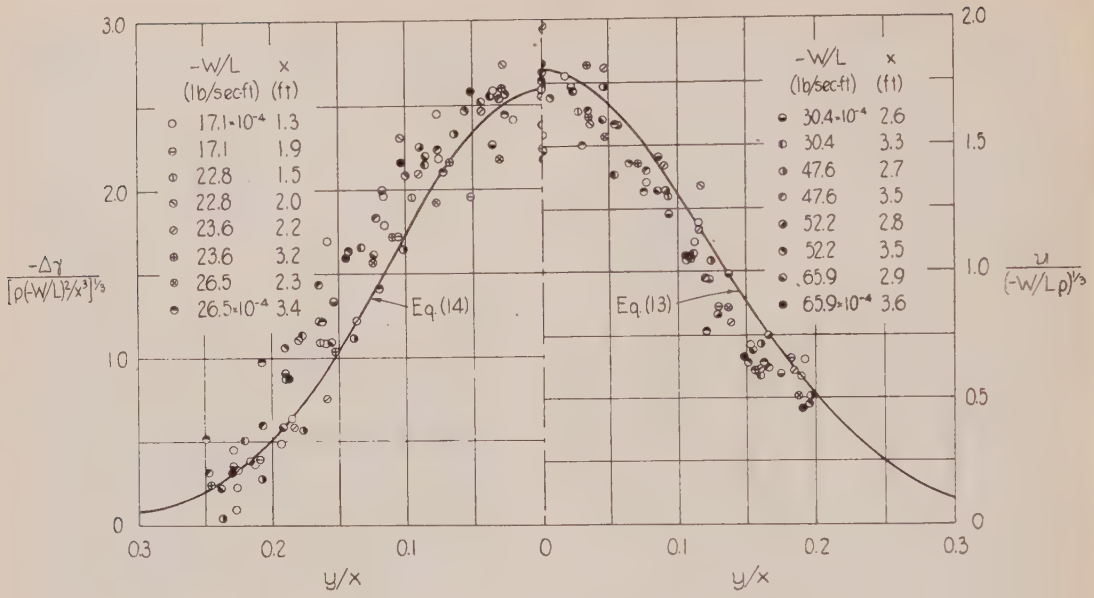


Fig. 2. Distribution functions for two-dimensional convection.

determination of the level at which, under ordinary circumstances, the initial laminar flow could be expected to become unstable. Because of the low rates of heat input required to yield readily measurable conditions of transition, it was necessary to use as the limiting heat source a burning cigarette (from a package carefully calibrated for rate of heat release), the smoke serving as a convenient tracing agent. The critical Reynolds number determined in this manner was of the form

$$\mathbf{R}_c = \frac{x_c}{\sqrt{\varrho\nu^3/(-W)}} \approx 10^5$$

in which ν is the kinematic viscosity. As can be determined herefrom, the critical height x_c was small in every case for which detailed measurements were made.

All velocity and temperature traverses were carried entirely across the convection zone, the plotted curves being essentially symmetrical about the maximum but often indicating a slight deviation of the flow from the vertical. In analyzing the results, the reference axis was adjusted to agree with the axis of symmetry. Upon reduction of data to the pertinent dimensionless forms, composite plots were made of velocity and weight distributions for both the two-dimensional and the axially symmetric cases, to verify the assumed similarity of flow at different levels and with different strengths of source and to determine the explicit distribution functions. For ease in comparison as well as economy of space, these plots are shown in half section: in Fig. 2 for two dimensions and in Fig. 3 for axial symmetry.

At once apparent is the appreciable scatter of points which seems to typify experimental studies of this nature — in part because of the difficulty of precise measurement and in part because of the sensitivity of such flow to slight disturbances. The scatter is not appreciably systematic, however, nor is it difficult to construct mean curves to indicate the distribution functions. The curves shown are normal-probability functions which best fit the data and at the same time satisfy Eqs. (4) and (5), and Eqs. (7) and (8), respectively. At least as a first approximation, there seems little doubt that the plotted curves describe the basic free-convection phenomenon. To be remarked, however, is the fact that use of the Gaussian function — not to mention the simplified analysis itself — loses significance as y/x or r/x becomes large.

If the probability curves plotted in Fig. 2 are assumed to represent the true distributions of u and $\Delta\gamma$, the corresponding expressions

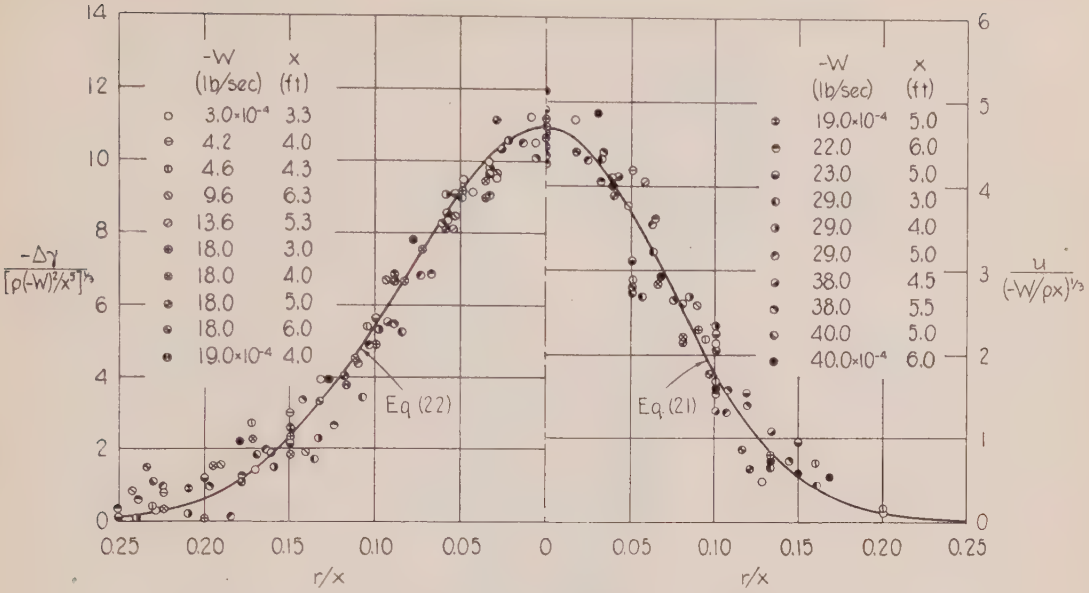


Fig. 3. Distribution functions for axially symmetric convection.

$$u = 1.80\left(\frac{-W/L}{\varrho}\right)^{1/3} exp\left(-32\frac{y^2}{x^2}\right) \quad (13)$$

$$\Delta\gamma = -2.6\left[\frac{\varrho(-W/L)^2}{x^3}\right]^{1/3} exp\left(-41\frac{y^2}{x^2}\right) \quad (14)$$

can be used conveniently to evaluate other characteristics of the mean pattern of motion. Thus, since $\psi = \int u dy$, the stream function is expressible implicitly as

$$\psi = 1.80\left(\frac{-W/L}{\varrho}\right)^{1/3} \int_0^y exp\left(-32\frac{y^2}{x^2}\right) dy \quad (15)$$

in which values of the integral are obtainable from probability tables. The lateral velocity of entrainment just beyond the diffusion zone, otherwise available from the condition that $v = -\partial\psi/\partial x$, is instead computed from the expression $2v_e = -d(Q/L)/dx$ as

$$v_e = \mp 0.28\left(\frac{-W/L}{\varrho}\right)^{1/3} \quad (16)$$

Specific relationships for Q, M, E, and F per unit length of source are obtained by integration as follows:

$$\frac{Q}{L} = 0.57\left(\frac{-W/L}{\varrho}\right)^{1/3} x \quad (17)$$

$$\frac{M}{L} = 0.72\left(\frac{-W}{L}\right)^{2/3} \varrho^{1/3} x \quad (18)$$

$$\frac{E}{L} = 0.53\frac{-W}{L} x \quad (19)$$

$$\frac{F}{L} = 0.72\left(\frac{-W}{L}\right)^{2/3} \varrho^{1/3} \quad (20)$$

For conditions of axial symmetry these expressions take the following forms:

$$u = 4.7\left(\frac{-W}{\varrho x}\right)^{1/3} exp\left(-96\frac{r^2}{x^2}\right) \quad (21)$$

$$\Delta\gamma = -11.0\left[\frac{\varrho(-W)^2}{x^5}\right]^{1/3} exp\left(-71\frac{r^2}{x^2}\right) \quad (22)$$

$$\psi = 0.024\left(\frac{-Wx^5}{\varrho}\right)^{1/3}\left[1 - exp\left(-96\frac{r^2}{x^2}\right)\right] \quad (23)$$

$$v_e = -0.041\left(\frac{-W}{\varrho}\right)^{1/3}\frac{x^{2/3}}{r} \quad (24)$$

$$Q = 0.153\left(\frac{-W}{\varrho}\right)^{1/3} x^{5/3} \quad (25)$$

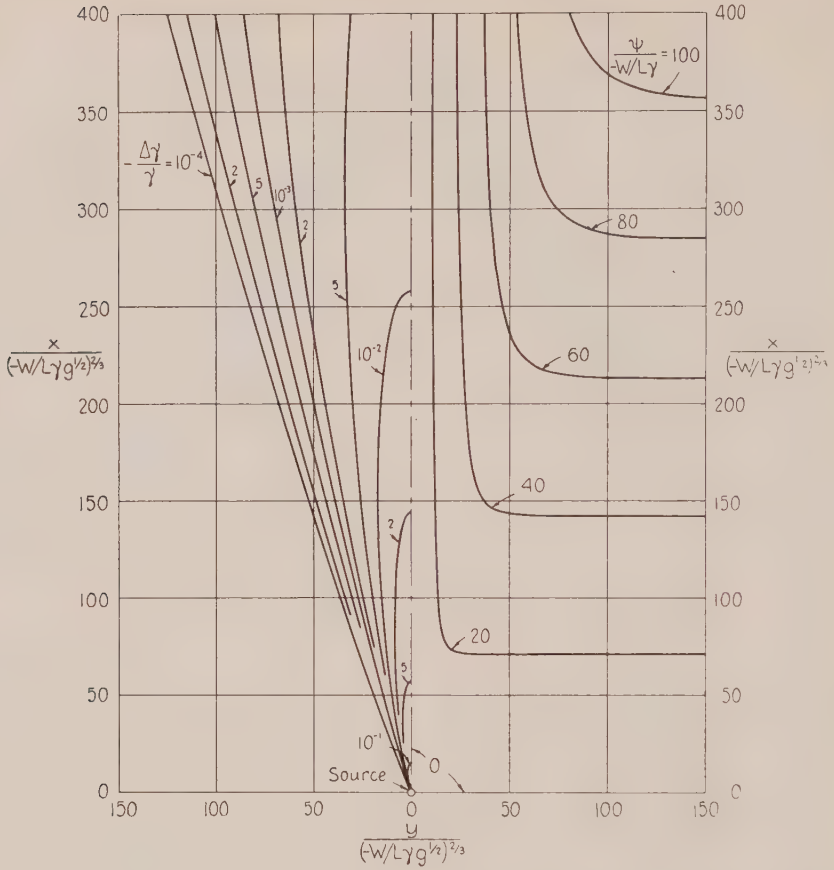


Fig. 4. Convection pattern over a line source.

$$M = 0.36\,(-W)^{2/3}\,\rho^{1/3}\,x^{4/3} \qquad (26)$$

$$E = 0.57\,(-W)\,x \qquad (27)$$

$$F = 0.49\,(-W)^{2/3}\,\rho^{1/3}\,x^{1/3} \qquad (28)$$

So far as the mean flow is concerned, it remains only to indicate in generalized coordinates the pattern of stream lines and lines of constant incremental weight for the two cases. This, however, requires three rather than two dimensionless groups of variables in each instance — the additional variable which is needed being a characteristic length to serve as a reference coordinate. Such a length is found in the quantity $(W/L\gamma g^{1/2})^{2/3}$ for two dimensions, and in the corresponding quantity $(W/\gamma g^{1/2})^{2/5}$ for axial symmetry. Upon introduction of these quantities, the dimensionless coordinate relationships for ψ and $\Delta\gamma$ become, for the line source,

$$\frac{\psi}{W/L\gamma} = \varphi_5\left[\frac{x}{(W/L\gamma g^{1/2})^{2/3}},\ \frac{\gamma}{(W/L\gamma g^{1/2})^{2/3}}\right]$$

$$\frac{\Delta\gamma}{\gamma} = \varphi_6\left[\frac{x}{(W/L\gamma g^{1/2})^{2/3}},\ \frac{\gamma}{(W/L\gamma g^{1/2})^{2/3}}\right]$$

and, for the point source,

$$\frac{\psi}{W/\gamma} = \varphi_7\left[\frac{x}{(W/\gamma g^{1/2})^{2/5}},\ \frac{r}{(W/\gamma g^{1/2})^{2/5}}\right]$$

$$\frac{\Delta\gamma}{\gamma} = \varphi_8\left[\frac{x}{(W/\gamma g^{1/2})^{2/5}},\ \frac{r}{(W/\gamma g^{1/2})^{2/5}}\right]$$

The corresponding plots of stream lines and lines of constant incremental weight are shown in half section in Figs. 4 and 5.

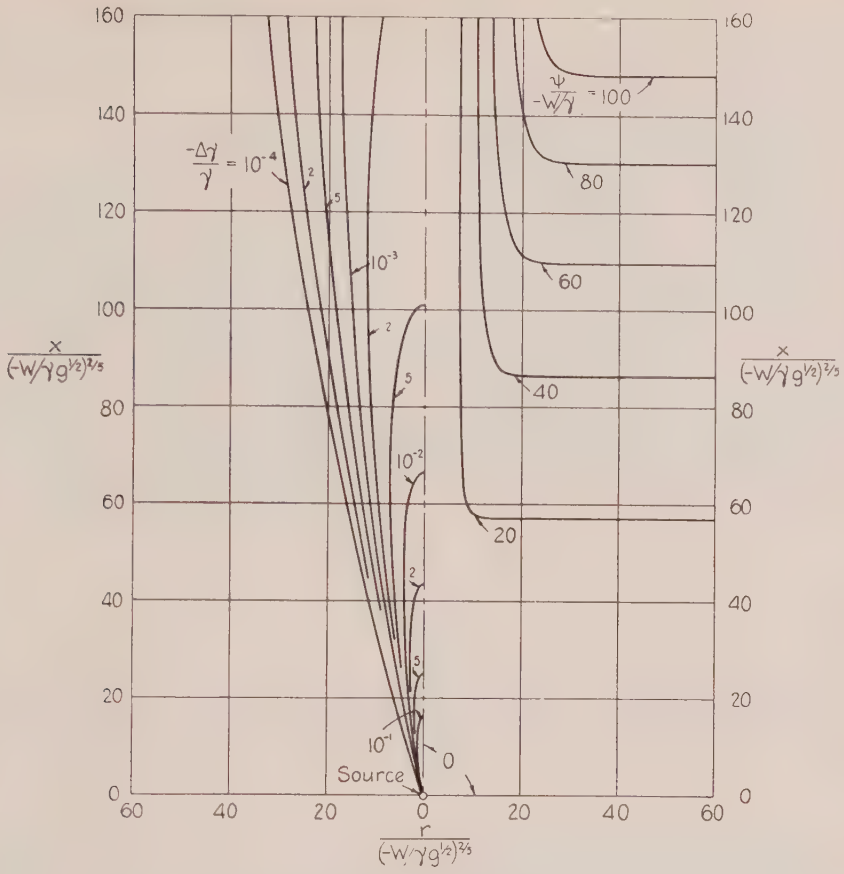


Fig. 5. Convection pattern over a point source.

As can be seen at once from these undistorted diagrams, the zones of pronounced vertical motion and change in weight density are concentrated about the area of symmetry and expand only slowly with elevation. Because of the fact that the measurements were restricted to these zones and that the analysis itself is not exact, the outlying portions of the stream-line patterns are purely schematic. Aside from this qualitative indication of the pattern of inflow, the diagrams are considered to represent with good approximation the basic characteristics of convection at any scale. They may readily be interpolated or extrapolated, as all curves in any diagram are geometrically similar by virtue of the dynamic similarity of the convection mechanism at successive elevations.

Conclusions

From the foregoing analytical and experimental studies of gravitational convection from both a line source and a point source, the following conclusions may be drawn:

(1) The motivating force in gravitational convection is the positive or negative buoyancy of the fluid, and the primary characteristics of the convection pattern can hence be analyzed without regard to the means whereby the buoyant effect is produced.

(2) Through use of an approximate method of analysis similar to that for boundary layers, wakes, and jets, it is possible to express as proportionalities the mean characteristics of gravitational convection for both

two-dimensional and axially symmetric turbulent motion.[1]

(3) Laboratory measurements of the mean convection patterns above boundary sources of heat have verified the approximate analysis and provided explicit relationships for the primary flow characteristics.

(4) Although both the analysis and the experiments represent a considerable simplification of the actual state of motion, the correlated results should permit evaluation of the role played by the primary convection process in more involved natural phenomena.

[1] Shortly before proof of the foregoing paper was received, there came to the authors' attention a German wartime publication on the same subject, the analysis proceeding from assumptions as to the secondary rather than the primary characteristics of the motion: SCHMIDT, W., 1941, Turbulente Ausbreitung eines Stromes erhitzter Luft, *Z. angew. Math. Mech.* **21**, 265—278, 351—363.

Acknowledgments

All phases of the investigation described herein were conducted in the laboratories of the Iowa Institute of Hydraulic Research under the direction of Hunter Rouse and were partially supported by the Office of Naval Research under Contract N8onr-500. The axially symmetric study was performed as a Ph. D. project by C. S. Yih, and the two-dimensional study as an M. S. project by H. W. Humphreys. Messrs. W. D. Baines and M. F. Andrews of the Institute staff rendered valuable assistance in the latter phase of the investigation.

Notes on the Surface Velocity Profile and Horizontal Shear across the Width of the Gulf Stream[1]

By WILLIAM S. VON ARX
Woods Hole Oceanographic Institution, Woods Hole, Massachusetts

(Manuscript received 1 August 1952)

Abstract

During a cruise across the Gulf Stream in October 1950 measurements of surface velocity were made both with the Loran-space-dead method and the electromagnetic method. A short account of the results is given with special reference to the velocity profile and the horizontal shear across the Gulf Stream.

The transverse surface velocity profile

In October, 1950, a short cruise was made across the Gulf Stream south of Cape Cod to observe the transverse surface velocity profile. Current velocities were observed by means of dynamic sections (Atlantic stations 4853 to 4865), the Loran space dead reckoning method, and the electromagnetic method (VON ARX, 1950). A preliminary reconnaissance of the area revealed an anticyclonic bend which was then crossed repeatedly at nearly right angles to the flow. Some of the results, shown in Figure 1, disagree among themselves but suggest the narrow width and sharply peaked form of profile which seems possibly to be characteristic.

Disagreement among observations stems partly from the average error of Loran, which at a range of several hundred kilometers from the transmitters amounts to about ± 1 kilometer. This amount of uncertainty in the position of each hydrographic station and of the drift of the ship while on station in the strongest part of the current, leads to a relative error

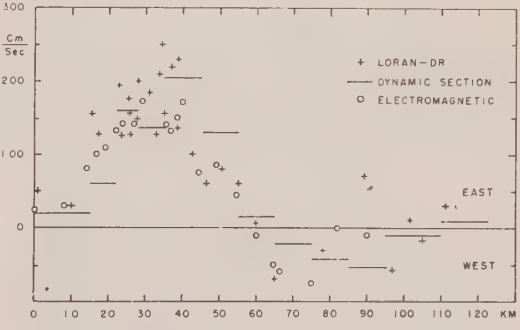


Fig. 1. Comparison of simultaneous measurements of surface velocity across the width of the Gulf Stream near the 69° W meridian, made by navigational, dynamic, and electromagnetic means. The area under a curve drawn through the eastward phase of each set of observations is 6.2, 6.0, and 5.4×10^8 cm²/sec., respectively. The dynamic observations were made and computed by Mr. L. V. Worthington. All observations made in October 1950 from R. V. ATLANTIS.

[1] Contribution No. 621 from the Woods Hole Oceanographic Institution. This work was supported by the Office of Naval Research under Contract No. N6onr-27701 (NR-083-004) with the Woods Hole Oceanographic Institution.

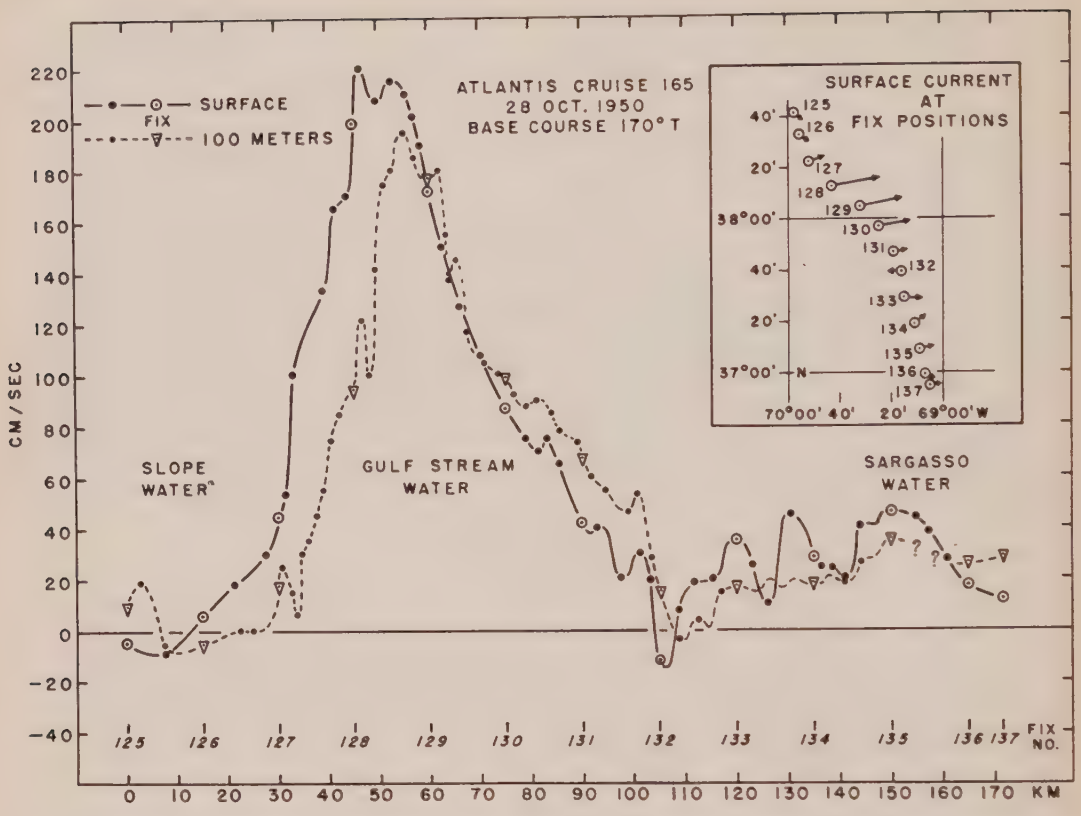


Fig. 2. Results of simultaneous electromagnetic measurements of the velocity at the surface and 100 meter levels. Open symbols denote points at which the direction of flow was measured. Closed symbols and the trend of the lines joining them denote the current speed at intermediate points assuming that the direction remained the same or changed uniformly between fixes of direction. The abscissa was taken at right angles to the direction of flow at fixes 128 and 129. The ordinates shown are 23 per cent less than the average of navigational measurements of current speed.

in the distance between stations 5 km apart perhaps as great as 20 or 30 percent. Similarly the errors of Loran and dead reckoning for navigational current measurements made on an hourly basis, may approach 50 cm/sec and become proportionally larger as the time interval is decreased. Measurements of surface current by the electromagnetic method are reduced in magnitude when the current structure extends downward through more than a negligible fraction of the total depth of water (Stommel 1948; Malkus and Stern 1952). In that the current structure of the Gulf Stream may extend a kilometer or more below the surface, and the loss of signal strength is roughly proportional to the ratio of the depth of quiet water below the current to the total depth, decrements as great as 40 percent are sometimes encountered in the Gulf Stream off the New England coast.

In the face of these uncertainties it has been necessary to counteract the limitations of one method with the advantages of another. In fine weather the evidence of navigation is a useful index of surface motion, but the method lacks resolving power. The electromagnetic method will indicate direction of flow and detailed changes in one component of motion, but requires calibration in magnitude. Both methods can be used underway.

Figure 2 shows the uncorrected results of electromagnetic sections made simultaneously at the surface and 100 meter levels. Navigational measurements made on the same traverse indicate that the electromagnetic measurements of surface current speed averaged 23

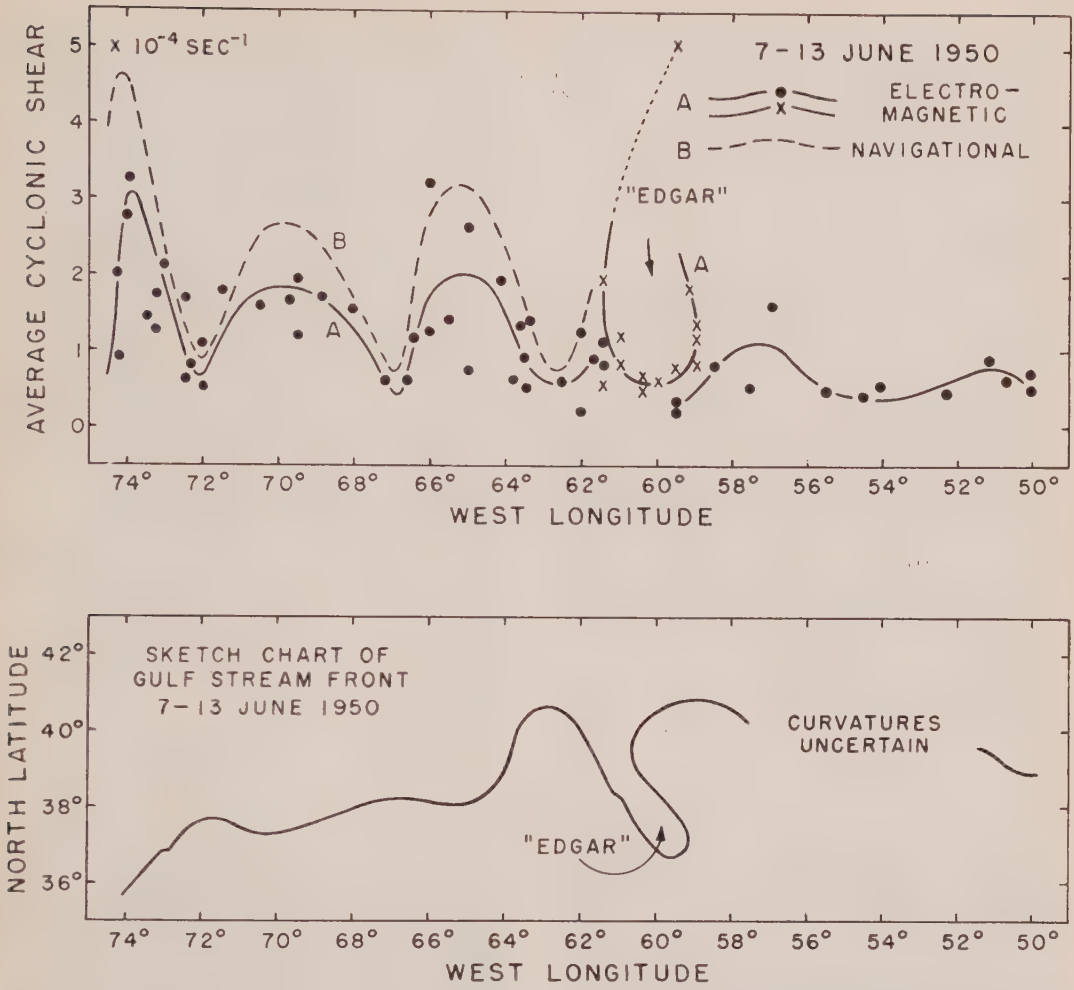


Fig. 3. Comparison of the intensity of cyclonic shear (between the Gulf Stream front and the current maximum) with the curvature of the front. Curves A and B in the upper frame represent the smoothed results of electromagnetic and navigational measurements of shear, respectively. The chart of the position of the Gulf Stream front in the lower frame was made from measurements of the greatest change of average temperature in the upper 200 meter layer. "Edgar" is a deep cyclonic bend that was observed to detach from the main current. Within this feature navigational measurements of shear were 1.7 times greater than electromagnetic measurements.

percent too small. Nevertheless a sharply peaked distribution of velocity across the current seems evident. The similar pattern of motion at the 100 meter level was displaced to the right of the surface profile along a seaward slope of approximately 1/80. The horizontal normal shear on either side of the current maximum was asymmetrical.

Horizontal shear

Data obtained during the multiple ship survey of the Gulf Stream of 1950 (FUGLISTER and WORTHINGTON, 1951) made possible a study of the change of the intensity of horizontal shear along the meandering current. The intensity of shear was first evaluated from a comparison of Loran and dead reckoning navigation, and then corrected for the local obliquity of the traverse to the direction of flow as indicated by electromagnetic measurements. In order that the curvature of flow might be neglected, the shear was evaluated and corrected for the obliquity of traverse to the direction of flow in increments of one or two hour's run.

It was found that the average intensity of horizontal shear normal to the flow on the right of the current maximum was equal to or less than the Coriolis parameter, while the average values of cyclonic shear on the left were equal to or greater than this. Such asymmetry of shear has been suggested from theory by HAURWITZ and PANOFSKY (1950).

These data from the multiple ship survey also show that the intensity of cyclonic shear on the left of the current maximum varies systematically with the sense of curvature of flow. Greater than average intensities seemed to be associated with cyclonic curvatures, while less than average intensities seemed to occur where the sense of curvature was anticyclonic. Due to the unsuitable distribution of observations, it was not possible to test for a similar correspondence of curvature with the intensity of anticyclonic shear on the right of the current maximum, or for the persistance of a seaward counter current.

In Figure 3 the upper frame contains the measurements of horizontal cyclonic shear, and the lower frame a schematic chart of the curvatures of the Gulf Stream front during the period when these observations were made. It can be seen from a comparison of the curves in the upper and lower frames that the intensity of shear does not seem to depend on the length of the radius of curvature of flow. The points through which curve A was drawn represent electromagnetic measurements of shear, and curve B represents the smoothed result of the navigational measurements. There is a systematic increase in the proportion of disagreement between the electromagnetic and navigational estimates of shear in the vicinity of cyclonic curvatures which suggests that the current extends to greater depths in the vicinity of cyclonic curvatures. Such an interpretation is consistent with an increase in cyclonic shear since vertical stretching and horizontal convergence of the flow in cyclonic curvatures would result in increased cyclonic vorticity on the left of the current maximum. There is some evidence that the current maximum itself shifts toward the left as well as that surface current speeds are a little higher in cyclonic curvatures of flow.

REFERENCES

FUGLISTER, F. C., and L. V. WORTHINGTON, 1951: Some results of a multiple ship survey of the Gulf Stream. *Tellus* **3**, 1—14.

HAURWITZ, B., and H. A. PANOFSKY, 1950: Stability and meandering of the Gulf Stream. *Trans. Am. Geophys. Union*, **31**, 723—731.

MALKUS, W. V. R., and M. E. STERN, 1952: Determination of ocean transports and velocities by electromagnetic effects. *Jour. Mar. Res.* (in press).

STOMMEL, HENRY, 1948: Theory of the electric field induced in deep ocean currents. *Jour. Mar. Res.* **7**, 386—392.

VON ARX, W. S., 1950: An electromagnetic method for measuring the velocities of ocean currents from a ship under way. *Papers in Physical Oceanography and Meteorology*, Vol. *XI*, 3: 1—62.

Composition of Atmospheric Precipitation

I. Nitrogen compounds

By ERIK ERIKSSON
Institute of Pedology and Institute of General and Inorganic Chemistry,
Royal Agricultural College of Sweden, Uppsala

(Manuscript received 13 August 1952)

Introduction

The purpose of this paper is to present a survey of published data on chemical analysis of atmospheric precipitation; to discuss different cycles proposed for inorganic compounds in atmospheric precipitation and, finally, to assess the importance of these compounds from different points of view. Investigations of rainwater with a view to determine atmospheric pollution in industrial areas will be mentioned incidentally but no results will be given in detail as they concern only small areas and abnormal conditions. The literature on this special subject is, furthermore, rather voluminous and should well deserve a separate review.

A. Historic review

The historic review presented below is largely built on existing reviews [see e. g. MILLER (213, 214), RUSSELL & RICHARD (254) and WAY (298)[1]] and will therefore not contribute anything new. The history of the research on chemical compounds in rain and snow is the history of the research on nitrogenous compounds in atmospheric precipitation. For a long time ammonia and nitrate nitrogen were the only compounds of interest while analysis for compounds like sulfate, alkali and alkaline earth metals have been carried out practically only for the last three decades. Chloride takes an intermediate position.

The first discovery of nitrate in rainwater was made by the German chemist MARGGRAF in the winter 1749—50 and in 1780—90 it was rediscovered by a Swede, TORBERN BERGMAN. In the beginning of the 1800's DE SAUSSURE in France showed the presence of ammonia in the atmosphere. Thus the two compounds of greatest interest were known to be present in the atmosphere at a time when the revolution in agricultural chemistry took place.

This revolution was mainly due to two other very important discoveries. In 1799 a Dutch, JOHAN INGENHOUSS, published his studies about the assimilation of CO_2 from air by plants and five years later DE SAUSSURE in France showed that inorganic constituents of plants were taken up by the plant roots from the soil. These two discoveries founded our present concepts of plant nutrition but it took, however, about half a century before they were accepted in the scientific world.

An important man in conveying these ideas

[1] References are listed in a bibliography to be published at the end of the second part of this study.

was no doubt JUSTUS VON LIEBIG although his role to a great extent was that of a propagandist. The actual outline of modern agricultural chemistry was made by another German, CARL SPRENGEL. At about the same time also BOUSSINGAULT in France, although leaning towards the old humus theory of plant nutrition, made a valuable contribution through some classical field experiments. Among other things he observed the difference between legumes and non-legumes with respect to nitrogen; the legumes were able to use atmospheric nitrogen for their nutrition. Despite of many years work he failed to find the reason for this difference.

BOUSSINGAULT was also the first to make analyses of rainwater in order to estimate the yearly contribution of combined nitrogen from rainwater.

In 1840 LIEBIG published his well-known "Die organische Chemie in ihre Anwending auf Agrikultur und Physiologie". In this work he strongly criticizes earlier concepts of plant nutrition and stated the modern concepts. In a later edition he changed his views on the uptake of nitrogen by plants. As he had found ammonia in plant sap he concluded that this compound, which was then known to be present in the atmosphere, was assimilated by the plants directly from the atmosphere in the same way as carbon dioxide. This was, of course, a rather obvious conclusion at that time. It may be of interest to cite him in a letter to "Revue Scientifique et Industrielle" given in English translation in the Farmer's Magazine (182). He states — "if the soil be suitable, if it contains a sufficient quantity of alkalies, phosphates and sulfates, nothing will be wanting; the plants will derive their ammonia from the air as they do carbonic acid. We know well that they are endowed with the faculty of assimilating these two aliments; and I really cannot see why we should search for their presence in the manures we use".—"Ammonia added with the manure may be useful but is certainly not necessary."

His quoted data on the amount of combined nitrogen which was brought down to the soil yearly with precipitation was very high, 27 kg/ha (1 ha = 10^4 m^2), and it was mainly this figure which initiated the many long-termed research projects on combined nitrogen in rainwater which were carried out during the later part of the 19. century. Among these the work carried out at Rothamsted is probably best known. (See LAWES & GILBERT (171), LAWES, GILBERT & WARINGTON (173, 174), MILLER (213), RUSSELL & RICHARD (254), WARINGTON (296), WAY (298, 299, 300). During the discussion which took place LIEBIG's theory about ammonia in plant nutrition was turned down on the basis of data from field experiments at Rothamsted [LAWES & GILBERT (171)], and his estimate of combined nitrogen in rainwater was shown to be too high as the yearly data from Rothamsted showed.

It is very probable that this controversy between LIEBIG and the Rothamsted scientists, combined with the increasing interest in the new concepts of plant nutrition, initiated many of the investigations which were carried out in Europe during this period. As will be shown below, similar investigations in U.S.A. were started much later.

Most of the inorganic compounds present in rainwater were known before 1850, but naturally, they did not arouse the same interests as nitrogen.

B. Nitrogen compounds in rainwater

1. Introduction

Most of the investigations carried out on rainwater concern nitrogen compounds, most frequently ammonia and nitrate, although a few cases also included nitrite and organically bound nitrogen, generally called albuminoid ammonia. Most of these investigations were carried out in the period 1870—1920. Fig. 1 shows the frequency of such investigations during successive 10-year periods, the frequency being given as the total number of years of investigation within each period. It will be seen from this figure that a maximum occurs in 1881—1890. The distribution for Europe is, further, approximatly normal while for other countries it is asymmetrical, starting with a relatively great number of years of investigations in the 1881—1890 period. In the period 1911—1920 there is a minimum in the number of investigations in Europe, probably connected with the occurrence of the first world war.

The contribution from Rothamsted is seen to take a central part in the distribution.

2. Nitrogen as ammonia and nitrate

a. Sampling and analytical errors

Before discussing the results some remarks on the sampling, the analytical methods and their errors will be made.

There are obviously two groups of errors which can be introduced; one group consists of errors due to the sampling technique and the storage of rainwater samples; the second group consists of errors in the analytical procedure used.

The sampling errors can be classified into inevitable errors and non-inevitable errors. Dust carried from the soil to the atmosphere (cyclic dust) and precipitated together with rain is an inevitable source of error as the dust particles may carry with them absorbed ammonia and nitrate or organic nitrogen compounds which are liable to give off their ammonia to the rainwater during the analytical procedure. The magnitude of this source may be considerable but is difficult to estimate. Existing data on yearly income of nitrogen compounds with the atmospheric precipitation may represent the total income but not the net amount, which from an agricultural point of view is of greater interest.

Cyclic dust seems to be the only inevitable error. Sampling errors which can be avoided include

a. Bird droppings.
b. Losses of ammonia from the sampling device due to aeration.
c. Transformation of the nitrogenous compounds in the sampling device by reactions with the sample and the material of the device or by microbiological activities during storage.

Bird droppings can produce errors in two ways, firstly by giving off nitrogen compounds as ammonia or as nitrate during the storage or the analytical procedure, and secondly by giving off substances which interfere with the analytical method used. In the second case it is, however, an error of the analytical method.

Bird droppings certainly can increase the content of nitrogenous compounds in rainwater. Johnson (154) compared protected and unprotected gauges and found that the bird droppings in the unprotected gauge raised

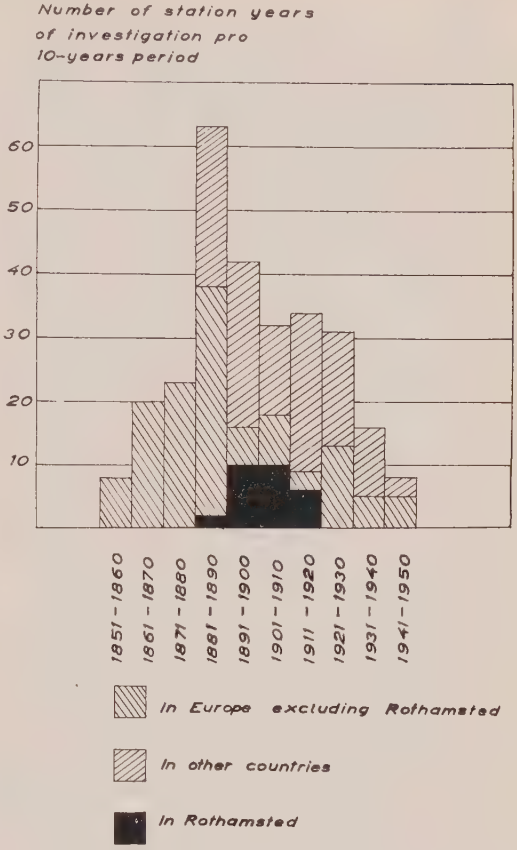


Fig. 1. Number of years of investigation (station-years) on nitrogen compounds in precipitation for successive 10-year periods.

the ammonia content considerably. The nitrate content was not materially affected.

Harrison & Williams (133) found definitely higher nitrogen contents when using copper funnels for sampling rainwater than when using glass funnels. They explained this as due to a greater contamination by bird droppings in the copper funnels. The birds found the copper funnels suitable as a resting place while they probably got scared by the light reflections in the glass funnels.

The error produced by bird droppings will, of course, depend upon the population density of birds at the place concerned and further upon the habits of the different species present. At seaside places the errors caused by birds may be very great. In the second paper by Miller (214) concerning investigations of rainwater from the west coast of Scotland several

samples were contaminated, apparently by bird droppings, to such a degree that the analyses and to be excluded. In inland places the errors due to bird droppings may not be great. A simple and effective device for preventing birds from using sampling gauges as resting places has recently been described.[1]

Losses of ammonia by aeration can be prevented by a suitable design of the collecting gauge. If the rainwater reacts with the material of the gauge the pH of the liquid will generally rise and thus enhance the ammonia losses.

Reactions between the sampling gauge and the rainwater have been studied to some extent. CROWTHER & STUART (82 *cf.* 81) found that copper containers were likely to neutralize rainwater, especially acid rainwater and in the above mentioned Swedish investigation it has been shown that galvanized iron in sampling gauges is liable to cause large errors due to reactions between Zn and water. As $Zn(OH)_2$ is formed, the pH will increase to such values that losses of ammonia can occur; further the hydrogen formed in this reaction is very active and reduces nitrate almost quantitatively to ammonia even when present in such low concentration as in rainwater.

Microbiological activities may occur during storage of rainwater. HANSEN (130) found that toluol was effective in preventing microbiological activity during storage. In many cases lead-lined gauges have been used. The small amounts of lead which go into solution are then sufficient to prevent microbiological activities during storage. This type of container was used in Rothamsted.

The analytical errors are due entirely to unsuitable methods for the analysis of ammonia and nitrate in rainwater. The most common methods for these compounds are colorimetric, using Nessler's reagent for the determination of ammonia, and phenoldisulphonic acid for the nitrate. The Nessler reagent is rather sensitive to the organic compounds always present in rainwater. If the determination is made directly in the rainwater, by adding the reagent, large errors can be produced, the magnitude and sign of which will depend on the special type of reagent which can be made up in a number of ways. ROMELL (251) has recently shown that water contaminated by pollen interferes with the Nessler reagent in the direct determination to a very high degree probably due to formaldehyde given off by the pollen grains. Many investigations have been carried out using this straight forward procedure and their results may be misleading. It is, further, probable that the errors are systematic and positive so that the given data are too high.

An indirect method using the Nessler reagent was used at Rothamsted. The ammonia in the rainwater was distilled off and the ammonia determined in the distilled portion by the reagent. This method, which has also been used in other places, gives quite reliable values. The ammonia in the distilled portion can also be determined by titration provided special care in taken during the distillation. This method has been used by HANSEN (130, 131) in Denmark and must be regarded as highly reliable.

The direct determination of nitrate by the phenol-disulphonic acid has been widely used but is not satisfactory as losses of nitrate may occur during the procedure (130). In older investigations of this type the nitrate figures may therefore be too low. In Rothamsted WARINGTON (296) introduced a new method of determining nitrate by reducing it to ammonia by a copper-zinc couple, followed by distillation and estimation of the ammonia by the Nessler reagent. Later Devarda's alloy was used for the reduction. This process seems to be quite satisfactory and has been used also in other places. The technique of using ion exchange resins in the sampling of atmospheric precipitation was described in the paper by Egnér and coll. cited in footnote on this page.

It is, of course, difficult to divide published rainwater nitrogen data into a reliable and an unreliable group based on the analytical methods used. In many cases no statement concerning the method is given, in other cases the data seem quite normal despite less satisfactory methods. There are so many factors involved (including the sampling factors) that rejection of data on the basis of unsatisfactory analytical methods cannot be done. The best way is to point out the most reliable investigations, basing the discussion mainly on these, leaving the rest as "probable" or in single cases as "quite improbable".

[1] EGNÉR, ERIKSSON and EMANUELSSON, 1949, Ann. Agr. Coll. Sweden *16*, 593—602.

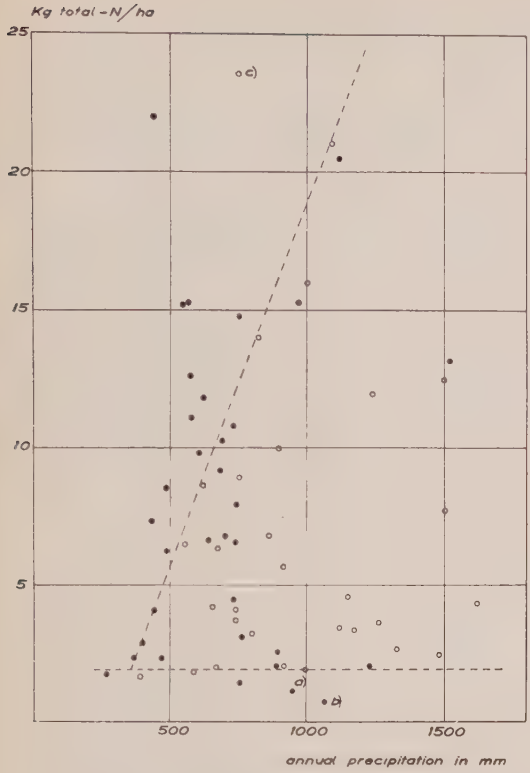


Fig. 2. The average yearly amounts of total nitrogen in precipitation plotted against the yearly precipitation.
Filled circles: European places.
Open » : non-European places.

b. Results

A survey of data up to 1900 has been given by MILLER (213). The investigations published since then do not change the general impression although they are quite numerous. For the sake of completeness existing data have been compiled; the averages for different periods and places being summarized in tables 1—3.

The results in tables 1—3 have been arranged diagrammatically in figs. 2 and 3 to show the distribution of yearly amounts as a function of the rainfall. In these figures a division between European and non-European places has been made.

The total nitrogen as ammonia and nitrate shows a very wide variation (fig. 2). There is a fairly marked minimum level of around 1.9 kg N/ha and year and it may be concluded that this is probably the smallest amount which is brought down by precipitation (except in deserts). The two exceptionally low values marked a) and b) in fig. 2 belong to coastal stations at Vifelstadir, Iceland, and Butt of Lewis in the Hebrides. It appears to be possible to draw another boundary for the distribution (the sloping line) which can be interpreted as the maximum nitrogen amount for a given yearly rainfall outside Europe. It is seen from fig. 2 that this maximum amount is greater in Europe than in other places. The places which lie above the sloping line are with one exception from Europe, namely from Russia, Germany (five from 1860—70 and one from 1928—35) and Denmark. The exceptionally high value c) comes from Buenos Aires. In the Danish investigation (HANSEN 129) the analyses were very accurate so there can be no analytical error involved, further there cannot be any great immediate contamination from industry.

The nitrate-nitrogen distribution is shown in fig. 3 and it will be seen that the minimum amount is not so distinct as in the case of total nitrogen. The sloping line of the maximum amount is, however, much more distinct than in fig. 2. On the other hand there is no marked difference between European and non-European places although there is a slight indication. The exceptionally high value a) is from an old German investigation (PROSKAU 1864).

It should be mentioned that the division of the data into a European and a non-European group has significance only from the point of density of population and industrial activities. Another possible division is tropical and non-tropical places; this is, however, of but little value as the investigations carried out in tropical places are rather few and incomplete. This can be seen in fig. 4, where the amount of total

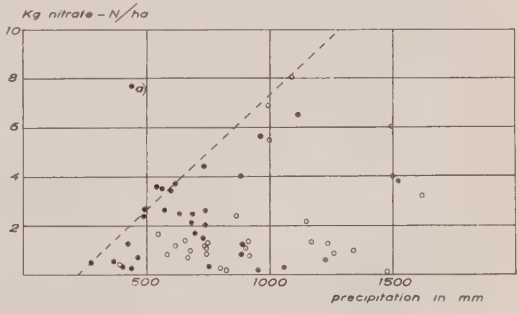


Fig. 3. The same as fig. 2 for nitrate nitrogen.

Table 1. Ammonia and nitrate nitrogen in precipitation in Europe

Ref.	Place	Period	Precip. in mm	mg N/l			kg N/ha		
				as H_3N	as NO_3	as tot. N	as H_3N	as NO_3	as tot. N
	Austria								
103	Feldkirch	1876	—	2.79	1.76	4.55	—	—	—
194	S:t Michael	1885—1887	1,114	1.17	.67	1.84	13.03	7.44	20.47
	Belgium								
235, 236	Gembloux	1889—1893	693	1.14	.35	1.49	7.92	2.39	10.31
146	Bulgaria		—	—	—	—	11.86	.96	12.82
	Denmark								
291	Copenhagen	1880/81—1884/85	577	1.91	.45	2.36	11.00	2.62	12.62
131	Askov	1922/23—1926/27	740	.71	.35	1.06	5.25	2.63	7.88
131	Blangsted	1925/26	600	1.07	.57	1.64	6.44	3.41	9.85
131	Spangsbjerg	1925/26	730	.88	.60	1.48	6.42	4.39	10.81
131	Hornum	1925/26	495	.74	.54	1.28	3.66	2.65	6.31
	Finland								
112	Storkyrö		—	—	—	—	1.8	—	—
	France								
176	Methroy	1877	760	.41	—	—	3.1	—	—
179—181	Montsouris	1876—1900	545	2.13	.66	2.79	11.60	3.60	15.20
	Germany								
213	Liebwerd, Bohemia	1877/78	620	1.3	.61	1.91	8.05	3.78	11.83
213	Peček »	1883/84, 1885/86	490	1.26	.50	1.76	6.17	2.45	8.62
53	Ida-Marienhütte	1865—1870	580	—	—	1.88	—	—	11.10
314	Kushen	1864/65—1865/66	369	.48	.16	.64	1.77	.59	2.34
314	Insterburg	1864/65—1865/66	642	.65	.39	1.04	4.15	2.49	6.64
314	Dahme	1865	427	1.42	.30	1.72	6.07	1.28	7.35
314	Regenwalde	1864/65—1865/66	568	2.08	.62	2.70	11.82	3.52	15.34
314	Proskau	1864/65	445	3.21	1.73	4.94	14.29	7.70	21.99
255	Weihenstephan	1928—1935	750	—	—	1.97	—	—	14.80
	Great Britain								
254	Rothamsted	1888—1916	730	.41	.20	.61	2.98	1.48	4.46
81	Garforth, Leeds	1906—1909	685	1.04	.32	1.36	7.09	2.16	9.25
214	Laudale, Ardgour	1907	2,250	.14	.06	.20	3.12	1.41	4.53
214	Barrahead, Berneray	1910—1913	895	.15	.14	.29	1.30	1.24	2.54
214	Shilay, Mon. Islands	1910—1913	1,230	.12	.05	.17	1.41	.66	2.07
214	Butt of Lewis, Stornoway	1908—1913	1,060	.04	.03	.07	.40	.34	.74
	Holland								
144	Groningen	1910—1912	700	.72	.24	.96	5.05	1.68	6.73
177	Hilversum (aver. of 6 places)	1932—1937	—	—	—	—	3.8	.7	4.5
	Iceland								
214	Vifelstadir	1911—1912	949	.09	.03	.12	.86	.28	1.16
	Italy								
24	Florens	1869—1875	970	1.00	.57	1.57	9.70	5.55	15.25
24	Vallombrosa	1872—1875	1,520	.62	.25	.87	9.40	3.80	13.20
232, 233	Scandicci	1888—1890	740	.61	.27	.88	4.55	2.00	6.55
21	Catania (outside town)	1889/89	466	.33	.16	.49	1.52	.75	2.27
	Norway								
252	Ås	1864/65	670	.31	—	—	2.1	—	—
51	Trondhjem	1928/29	889	.13	.10	.26	1.17	.87	2.04
51	Voll (7km east of Trondhjem)	1928/29	758	.14	.05	.19	1.06	.38	1.44
	Russia								
239	Odessa	1903	275	.49	.18	.67	1.35	.50	1.85
292	Novosil	~ 1905	—	—	—	—	—	—	7.50
164	Taschkent	1945/46—1946/47	~ 400	.6	.1	.7	2.5	.4	2.9
301, 302	Ploty	1900—1903	445	.85	.06	.91	3.88	.27	4.15
	Sweden								
106	Flahult	1909	827	.45	.18	.63	3.72	1.46	5.18
	Switzerland								
120—124	Basel	1870/71	890	—	.46	—	—	4.05	—

Table 2. Ammonia and nitrate nitrogen in precipitation in Asia, Australia, New Zealand and Africa

Ref.	Place	Period	Precip. in mm	mg N/l			kg N/ha		
				as H_3N	as NO_3	as tot. N	as H_3N	as NO_3	as tot. N
	Asia								
	East Indies								
200	Pasoeroean, Java	1891/92	920	.14	.09	.23	1.27	.79	2.06
250	Sumatra, Deli	1925—1940	2,045	—	—	1.5	—	—	30.4
80	Malaya	1941	—	—	—	—	—	.10	—
	India								
243	Allahabad	1932	—	.50	.88	1.38	—	—	—
213	Calcutta	1891	1,170	.17	.12	.29	2.00	1.34	3.34
137, 175	Cawnpore	1904/05	1,260	.22	.06	.28	2.73	.85	3.58
137, 175	Dehra Dun	1904/05	2,220	—	—	.18	—	—	4.02
213	Madras	1888—1893	995	—	—	.19	—	—	1.91
86	Sylhet	1931/32	3,950	.13	.09	.22	5.06	3.74	8.80
17, 60	Hanoi, Tonkin	1902—1909	1,650	.79	2.94	3.73	13.0	48.5	61.5
18	Ceylon,	1898/99	2,090	.20	.07	.27	4.09	1.43	5.52
162	» Peredeniya	1941/42—1942/43	2,850	.27	.18	.45	7.82	5.22	13.04
	Japan								
157	Tokio	1883/84	1,337	.13	.07	.20	1.67	.97	2.64
216	»	1938	—	.58	—	—	—	—	—
216	Kobe	1938	—	.28	—	—	—	—	—
216	Hamamatu	1938	—	.19	—	—	—	—	—
	Australia								
56	Brisbane, Queensland	1908	1,150	.22	.19	.41	2.48	2.14	4.62
56	Roma, »	1908	660	.42	.21	.63	2.77	1.37	4.14
	New Zealand								
125	Lincoln, Canterbury	1884—1887	755	.08	.17	.25	.56	1.27	1.83
	Africa								
156	Bloemfontein, South Africa	1910/11—1911/12	550	.87	.30	1.17	4.76	1.67	6.43
156	Cedera » »	1912	680	.78	.14	.92	5.27	.97	6.24
156	Durban » »	1911—1912	910	.47	.15	.62	4.23	1.39	5.62
156	Grahamstown, » »	1911/12	590	.18	.14	.32	1.05	.82	1.87
156	Kokstad » »	1912	670	.19	.11	.30	1.25	.75	2.00
149	Pretoria » »	1904/05	620	1.19	.20	1.39	7.40	1.21	8.61
44	Mauritius	1895	(1,500)	(.43)	(.40)	(.83)	6.45	6.00	12.45
191	»	1938	1,002	3.38	—	—	3.80	—	—
222	Reunion	1886—1887	(1,000)	—	.69	—	—	(6.9)	—

nitrogen and nitrate nitrogen is plotted against the latitude of the place in question. It will be seen that the tropical zone is so poorly represented that it is difficult to make a fair comparison between this zone and the temperate ones. Fig. 4 is interesting, however, as it shows a very high frequence of large total nitrogen values in the northern hemisphere. The nitrate nitrogen seems to be fairly evenly distributed throughout all the zones.

Returning to figs. 2 and 3 the sloping border lines suggest that the variation in maximum nitrogen amount tends to be proportional to the amount of precipitation, or, expressed in a different way, that the maximum concentration of nitrogen is independent of the annual amount of rainfall. The significant difference between European and non-European places is that the maximum nitrogen concentration in European rainfall is able to reach somewhat higher values than in non-European. For nitrate nitrogen there is no such pronounced difference (fig. 3).

As the nitrate nitrogen does not show any pronounced difference in the distribution between European and non-European places the difference in fig. 2 must be due to the ammonia nitrogen; the ammonia nitrogen

Table 3. Ammonia and nitrate nitrogen in precipitation in America

Ref.	Place	Period	Precip. in mm	mg N/l			kg N/ha		
				as H_3N	as NO_3	as tot. N	as H_3N	as NO_3	as tot. N
	North America								
271—275	Ottawa, Canada	1908—1924	860	.51	.28	.80	4.43	2.39	6.82
310	Alfred, N.Y. U.S.A.	1923/24—1924/25	820	1.67	.03	1.70	13.70	.28	13.98
310	Brockport » »	1923/24—1925/26	800	.36	.03	.39	2.91	.26	3.17
75	Geneva » »	1919—1928	900	.99	.12	1.11	8.91	1.07	9.98
332, 335	Ithaca » »	1915/16—1925/26	750	1.07	.12	1.19	8.00	.91	8.91
110	Goodwell, Oklahoma U.S.A.	1930	390	.28	.12	.40	1.11	.48	1.59
115	Kentucky, six places U.S.A.	1922/23	1,090	1.19	.74	1.93	13.00	8.02	21.02
104	Manhattan, Kansas U.S.A.	1886/87—1889/90	740	.39	.16	.55	2.87	1.16	4.03
15, 115, 117, 160, 161, 234, 246, 269, 290, 305, 306, 313,	Mont Vernon, Iowa U.S.A. (8 years)	1912/13—1931/32	740	.35	.15	.50	2.63	1.08	3.71
102	Salt Lake City, Utah U.S.A.	1891—1893	—	—	—	—	5.65	.40	6.05
289	Missouri, U.S.A.	1894—1895	1,120	.24	.07	.31	2.63	.82	3.45
	West Indies								
133	Barbados	1886—1889	1,620	.07	.20	.27	1.12	3.24	4.36
61	Trinidad	1901—1912	1,480	.16	.01	.17	2.37	.10	2.47
	South America								
134	Georgetown, Br. Guiyana	1890—1909	2,520	.05	.08	.13	1.13	2.06	3.19
196, 222	Caracas, Venezuela	1889/90, 1895	(1,000)	1.55	.55	2.10	(15.5)	(5.5)	(21.0)
268	Montevideo, Uruguay	1912	1,504	.24	.27	.51	3.68	4.03	7.71
58	Buenos Aires, Argentine	1906	750	2.96	.18	3.14	22.2	1.33	23.53

content varies more in Europe than in other places.

The proportions of ammonia to nitrate nitrogen is shown in fig. 5, arranged in the same way as in fig. 4. It has been concluded by MILLER (213), that trophical rains contain a higher percentage nitrate nitrogen than non-tropical. As seen in fig. 5 this seems to be true when comparing zones II and III and is probably due to the earlier mentioned greater concentration of ammonia in European rainfall. Comparing zones III and IV it seems, however, that there is no reason to believe the proportions to be really different. As a matter of fact there are only two places in the tropical zone that show a low ammonia percentage.

In this connection it should be mentioned that some investigations show an abnormally high ammonia percentage, especielly those carried out in the state of New York, U.S.A. There is probably some error involved, either in the sampling or the analysis which have led to a reduction of nitrate to ammonia, or in the analysis, in which contamination present has affected the ammonia figures. The second explanation is more probable. It has been pointed out earlier by ROMELL (251), that the ammonia figures from these places are abnormally high and he suggested interferences in the direct determination of ammonia by Nessler's reagent, caused by substances given off by pollen grains in the rainwater samples.

It is difficult and probably not justified to work out an average of nitrogen brought to the soil by rainwater in different parts of the world as the variation between different places probably is more significant than an average.

c. Monthly variation

The amount of combined nitrogen brought to the soil with the rain each month will vary not only with the rainfall but also with the general climatic conditions and the habits and occupation of the population. For some of the

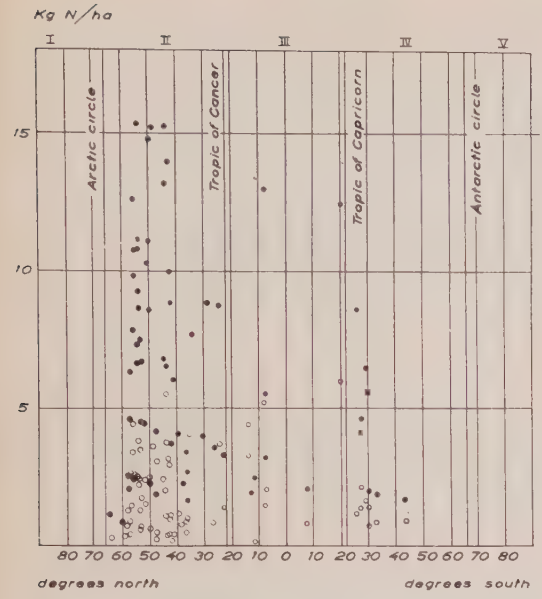


Fig. 4. The average yearly amounts of total nitrogen and nitrate nitrogen for different places plotted against their latitude.
Filled circles: total nitrogen.
Open » : nitrate »

places recorded in tables 1—3 monthly averages are given and a few representative investigations of these are shown in table 4.

In Gembloux, a place where atmospheric pollution is fairly high, the monthly amount of nitrogen is fairly well correlated with the monthly amount of rainfall. The season has, however, also a direct influence on the amounts. This can already be seen from the figures in the table and is clearly shown by statistical analysis. Assuming that the relationship between amount and precipitation is linear and the seasonal influence is a simple periodic one, the relation arrived at reads

$$N = 0.274 \cos 30\,(T + 0.3) + 0.0143\,p - - 0.166 \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{(A)}$$

where N is the amount of ammonia nitrogen in kg/ha/month, T is the number of the month starting with January equal to zero and p is the monthly rainfall in mm. The argument of the cosine function is expressed in degrees. Thus, there is a higher ammonia content in the precipitation in winter than in summer which may be due to an increased coal combustion during the winter.

In Rothamsted there is a maximum in the amount of ammonia nitrogen in August and a minimum in February. The influence of the rainfall is obscured by the fact that it also varies periodically showing a maximum in October. The ammonia concentration during different months (weighed averages, not given in the table) shows, however, two maxima, one in the spring and one in the late summer as a generally somewhat higher level during the warm season than during the cold.

A significant feature for Ottawa is that the rainfall varies very little from month to month. The variation in the amounts of ammonia nitrogen can therefore be ascribed mainly to seasonal influences. An analysis of the data gives a relation which reads

$$N = 0.150 \sin 30\,(T - 4.1) + 0.099 \sin 60\,(T - 1.5) + 0.368 \ldots\ldots\ldots\ldots\ldots\ldots \text{(B)}$$

and is graphically illustrated in fig. 6. The seasonal variation has thus been resolved into two components, the first of which follows the expected yearly temperature of the air at such a place. The second one has two periods within the year with maxima in spring and autumn just as indicated by the figures on the ammonia content in Rothamsted. Both of the

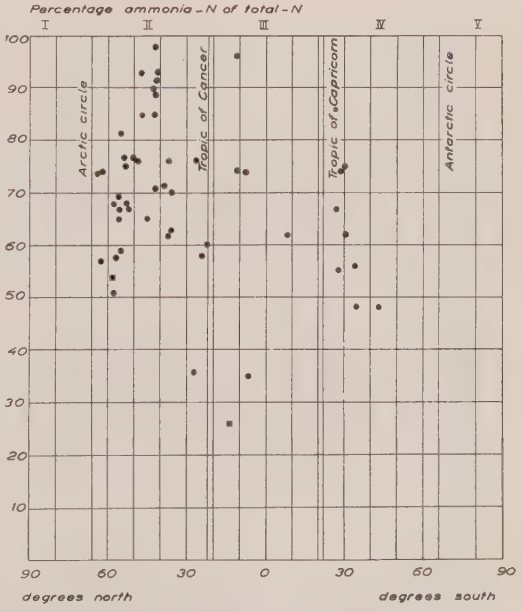


Fig. 5. The percentage ammonia nitrogen of total for different places, plotted against their latitude.

Table 4. Monthly distribution of amounts of nitrogen compounds in precipitation in a few representative places

Ref.	Place and item	Jan.	Febr.	March	April	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.
235	Gembloux												
	H_3N—N, kg/ha...	.54	.23	.79	.57	.53	.81	.81	.97	.42	.89	.76	.61
	NO_3—N, kg/ha...	.19	.03	.15	.18	.32	.26	.29	.33	.17	.23	.15	.11
	rainfall, mm.....	44	13	51	46	59	93	104	85	45	62	52	42
254	Rothamsted												
	H_3N—N, kg/ha...	.21	.19	.23	.23	.26	.27	.28	.31	.24	.26	.25	.25
	NO_3—N, kg/ha...	.10	.10	.11	.11	.13	.13	.12	.15	.12	.18	.12	.13
	rainfall, mm	55	51	61	43	51	64	61	72	48	82	67	78
275	Ottawa												
	H_3N—N, kg/ha...	.21	.15	.22	.43	.43	.40	.33	.53	.51	.61	.35	.25
	NO_3—N, kg/ha...	.14	.17	.17	.26	.19	.22	.28	.31	.17	.20	.14	.13
	rainfall, mm	81	64	72	65	74	72	70	74	68	80	69	71
133	Barbados												
	H_3N—N, kg/ha...	.06	.05	.05	.05	.06	.08	.17	.23	.12	.10	.11	.06
	NO_3—N, kg/ha...	.12	.10	.10	.10	.21	.18	.30	.40	.21	.23	.50	.31
	rainfall, mm	76	52	42	30	93	114	177	257	203	141	208	117
133	British Guiyana												
	H_3N—N, kg/ha...	.10	.16	.18	.19	.20	.11	.11	.19	.01	.07	.17	.09
	NO_3—N, kg/ha...	.33	.26	.12	.18	.27	.34	.32	.15	.08	.06	.17	.19
	rainfall, mm	336	259	184	205	312	328	278	173	64	51	142	380
162	Peredeniya												
	H_3N—N, kg/ha...	.57	.11	.75	.26	.46	.36	.21	.87	1.75	.55	1.23	.70
	NO_3—N, kg/ha...	.16	.34	1.09	.29	.88	.58	.15	.30	.43	.33	.54	.13
	rainfall, mm	64	45	146	279	603	296	112	183	203	349	359	226

sine functions in relation (B) are statistically significant. It would be of interest to see if this distribution of ammonia nitrogen might be related to the occurrence of forest fires.

In the tropical places Barbados, British Güinea and Peredeniya the variation in rainfall is very great and obscures the direct influence of the season. A correlation between the amounts and the rainfall is indicated by the data.

The nitrate nitrogen does, in general, show less influence of the season than ammonia but is fairly well correlated with the rainfall. From the Ottawa figures it can be seen, however, that there is a certain seasonal influence. A relation of type (B) for nitrate nitrogen reads

$$N = 0.070 \sin 30\,(T - 2.6) + 0.022 \sin 60\,(T + 0.3) + 0.199 \quad \text{(C)}$$

both sine functions being statistically significant although the contribution from the second one is rather small. The maximum due to the first component occurs between June and July; it differs from the ammonia relation in that respect by one and a half month. The maxima of the second component occurs in February and August. Its significance from meteorological point of view is doubtful.

d. Yearly variation and trends

The variation of the yearly amounts of nitrogen compounds at a certain place is in general fairly large. This may be due to a number of factors such as yearly rainfall, prevailing winds and the habits and occupation of the population dwelling in or around the place in question. These factors vary periodically but some may have a very long period and give rise to what is called yearly trends. A closer study of these trends necessitates long-termed series of observation, especially when the yearly variation is large.

At two places, namely at Rothamsted and in Ottawa, the yearly amounts of ammonia

and nitrate nitrogen in precipitation have been estimated for relatively long periods. In Ottawa (from 1908 to 1924) the yearly amounts have a tendency to increase with time. This increase is especially noticeable for the last four years. The average amount of ammonia nitrogen for the first ten years is 3.88 kg N/ha while the average for the whole period (17 years) is a high as 4.40 kg N/ha. The increase is probably due to increasing human activities. For nitrate nitrogen no trend at all is detectable.

At Rothamsted (from 1888 to 1916) the amounts of ammonia does not show any clear trend but the nitrate nitrogen, as was also pointed out by RUSSEL and RICHARD (254) increases slowley but steadily with time. The reason for this increase may be difficult to find.

In the period 1908 to 1916 the two series overlap. There is, however, no obvious correlation neither between the amounts of ammonia nor between the amounts of nitrate at the two places. This is also interesting as it shows that the factors reponsible for the yearly variation may be fairly local in extension.

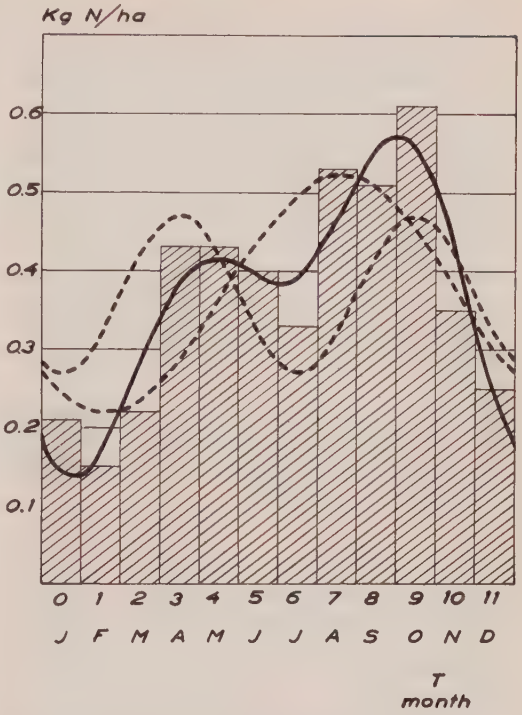


Fig. 6. The monthly amounts of ammonia nitrogen of Ottawa precipitation.
Dashed lines: the two components of the full drawn line mentioned in the text.

3. Albuminoid nitrogen

This name has been assigned to the ammonia released from organic matter in the rainwater when treating it with oxidizing agents, for instance $KMnO_4$, or by Kjeldahl digestion. There are a few data available for albuminoid nitrogen in rainwater. In Mont Vernon, Iowa, (ref. 15, 114, 115, 160, 161, 234, 246, 269, 290, 305, 306, and 313) it is estimated to be 2.7 kg/ha and year as an average for 9 years and at an average rainfall of 755 mm. In Ottawa, Canada, (271—275) the yearly amount was 0.98 kg/ha as an average for 17 years and a rainfall of 860 mm. At Dehra Dun in India (86), the amount for 1931—32 was 6.05 kg/ha at 2,950 mm rainfall. In Rothamsted MILLER (213) has estimated the amount to 1.5 kg/ha/year, the estimate being based on 69 analyses. In New Zealand GRAY (125) found 0,5 kg/ha.

From these scattered data it is seen that the contribution may be considerable in some places. Some of this nitrogen is apparently derived from cyclic dust and, thus, does not contribute anything to the nitrogen economy of the soil.

4. Nitrite nitrogen

The nitrite nitrogen, when reported, is in general very low. In some investigations it is automatically included in the nitrate fraction as the determination of nitrate is made by reducing it to ammonia by a metal alloy. Such a reduction obviously also includes the nitrite.

FAILER and WILLARD (104) tested daily samples of rain in Kansas qualitatively for nitrite. Of 171 samples 75 per cent gave a clear positive reaction, 8 per cent no reaction and the remaining 17 per cent a doubtful reaction for nitrite. The frequency of nitrite present was higher during summer than during winter.

The investigations carried out in Mont Vernon, Iowa, show variations in the daily samples from a trace to 0.051 mg nitrite nitrogen/l or roughly about 10 per cent of the nitrate nitrogen. The same was also esti-

mated at Sylhet, India (86) where the total nitrite nitrogen for a year amounted to 0.46 kg/ha, most of it being delivered during April—June. Analysis of rainfalls near Bombay over a half year period gave nitrite nitrogen contents ranging from traces to 0.002 mg N/l. In Japan (216) the average contents for three places were 0.0027, 0.0040 and 0.0067 mg N/l as nitrite. An old French investigation (69) show comparatively high values, but from the more recent ones quoted here it seems likely that the amounts are relatively low. Some investigations carried out in Australia will be discussed later.

Most of the data quoted here were obtained on samples from single rainfalls, analysed very soon after the precipitation. Hence the figures should be quite reliable.

5. Nitrogen compounds in fog, dew, frost and condensed water from the air

The compositions of these forms of atmospheric precipitation have been investigated to some extent.

a. Fog. In 1853 Boussingault (50) found more ammonia and nitrate in fog than in rain. The amount of ammonia ranged from 3—9 mg N/l. This was later confirmed by Levy (180) who, however, found a low nitrate content in fog. In Belgium Peterman and Graftiau (236) made three analyses which gave 2.60, 5.64 and 5.00 mg total-N/l and in Bologna in Italy Casoli (62) found values ranging from 14—58 mg ammonia nitrogen/l. From Russia it is reported (244) that fog, dew and frost contained two and a half times as much ammonia as did rain and two and a half times less nitrate.

b. Dew. Boussingault (50) found as much ammonia in dew as in fog but much less nitrate. In Cawnpore in India (137, 175) the analysis of dew gave 0.85—2.65 mg ammonia nitrogen and 0.51—4.12 mg nitrate nitrogen/l of dew. For the six month period in 1904—05 the amount of ammonia and nitrate nitrogen delivered to the soil by dew amounted to 0.06 kg/ha of each. Also Levy (180) reported that dew was rich in ammonia.

Recently Quitmann and Cauer (241) compared the composition of condensation water near the ground with that of rain on the same occasion. They got (two cases)

	H_3N mg/l	NO_2 mg/l
Condensation water .	10.0, 10.0	0.10. 0.05
Rain	1.5, 1.0	0.35, 0.10

which shows a considerable difference in composition between these two forms of precipitation.

c. Frost. The Belgium (236) investigation gives an average of 7.52 mg total nitrogen per litre for five analyses; in Bologna (62) the range of concentration of ammonia in frost was 22.4 to 44.7 mg N/l. Some recent analyses from Germany of Mrose (218) show lower values with a variation ranging from 0—4.8 mg ammonia nitrogen and 0—3.7 mg nitrate notrogen/l.

From all these results it can be concluded that those forms of precipitate which are derived from near the soil surface as a rule are much richer in ammonia than ordinary rainwater.

In this connection some observations on the rate of rainfall and size of raindrops on the ammonia content of rain should be mentioned. As early as 1870 Rosing (252) in Norway observed that a very fine rain (drizzle) had a higher ammonia content than ordinary rain. Later also Hansen (131) pointed out that a long continous rain falling at a low rate had a higher ammonia content than showers.

6. Origin of ammonia nitrogen in rain

a. *The equilibrium in the air—sea water system*

In connection with his fundamental plant nutrition experiments Boussingault suggested that the sea was the source for ammonia in air. His theory was strongly supported by Schlösing who in a series of papers (257—265) discussed this possibility on the basis of his own experimental data. In an introductory paper he outlined the theory as follows: As all plants take up nitrogen only in combined form (nitrogen fixing bacteria were not known that time) and some loss of combined nitrogen occur during the decay processes of plants and animals, the total reserve of combined nitrogen in air must decrease unless there are some processes by which it is brought to the air or converted to combined form. It is known that electric discharges in the air oxidize nitrogen

to nitrate which is carried down in rainwater. This amount can, however, not account for the nitrogen taken up by the plants and lost in drain water; there must exist a big reservoir which can deliver combined nitrogen to the air. This reservoir is the sea which contains relatively large amounts of ammonia. The ammonia cycle is then this: The nitrate in drainage water is carried to the sea where it becomes reduced to ammonia which diffuses into the air and is either absorbed from the air by the soil or even the plants or is washed down by rain. The losses which occur (free nitrogen) during this cycle are compensated by formation of nitrate in electrical discharges.

SCHLÖSING tried to verify his theories experimentally. He determined the equilibrium constant in the air and at different temperatures. Further he analysed air for ammonia and from his results he draw the conclusion that his theory was correct. He also studied the absorption of ammonia by the soil and estimated that about 40 kg ammonia nitrogen was absorbed from air per year and hectare. Further he found that rainfall did not always make the ammonia content of air to decrease. Sometime it even increased after a rainfall all depending on how the equilibrium conditions in the place where the rain was formed differed from the equilibrium conditions near the surface.

His theory was, however, not accepted everywhere. An investigation from Spain (10) indicated that the ammonia content in rain was in general lower when the wind came from the sea and the same was observed also by POSNJAKOV (239) in Russia. In Great Britain MILLER (214) started an investigation on rainwater collected at a number of seaside places and isles on the Scottish west coast and found in general low ammonia contents (see table 1), Similar results have been obtained later by BRAADLIE (51), HANSEN (131), and LEEFLANG (177). Unless the higher inland figures are due to a greater portion of cyclic dust present it is obvious that these observations do not tally with SCHLÖSING's theory about the sea as a source for ammonia in air.

The actual ammonia distribution in the sea is fairly well-known. The concentration at the surface in the temperate and tropical zones has been estimated to 0.03 mg ammonia/l. (WATTENBERG (297) see also BUCH (57). It decreases downwards to around 0.01 mg at 100 m depth and to around 0.005 mg at 200 m depth. Already this gradient could suggest that there is a steady absorption of ammonia from the air by the sea. But the gradient is normally explained in a different way. The nitrate shows an opposite gradient. It is practically zero at the surface but increases downwards to reach a comparatively constant value below 200 m depth. This is due to the content of plankton which is greatest at the surface but which is limited there by the nitrate supply, in this way keeping a relatively high ammonia concentration at the surface. At greater depths the limiting growth factor is light; the nitrate level is therefore greater and that of the ammonia less. This organic life thus acts as a catalyst for the photo-chemical reduction of nitrate to ammonia, the equilibrium being determined by the light intensity. In polar regions and elsewhere where vertical convection takes place no such gradient is observed with the well-known result of an intense plankton production.

From this point of view the concentration gradient of ammonia in the sea does not necessarily indicate a transport of ammonia from the air to the sea.

Calculations by BUCH (57) show on the other hand that at equilibrium with the ammonia content in sea water the concentration of ammonia in air should be only 0.4 $\mu g/m^3$ air. BUCH, however, points outt hat reliable figures on the ammonia pressure at equilibrium have been obtained only at relatively high ammonia concentrations. An investigation at lower concentration would be desirable.

The actual ammonia concentration in air has been determined in a few cases, mostly long time ago. The most reliable results are summarized in table 5. Of these investigations that of LEVY from Montsouris covers a whole year with daily analyses. It is seen that the monthly variation is small and there was no clear trend between different months. In Budapest the values were higher in summer than in winter. The results from Buenos Aires are comparatively high probably due to atmospheric contamination. The figures from Pic du Midi (at 2,900 m altitude) are probably the most realiable ones existing for the ammonia content in uncontaminated air. The average of these is 13.5 $\mu g/m^3$.

Table 5. Ammonia in air

Ref.	Author and year	Place	μg H_3N/m^3	Remarks
179	Levy 1880	Montsouris, France	21—29	
113	Fodor 1881	Budapest, Hungary	39	average
219	Müntz et al, 1882	Pic du Midi, summit, France	7.2—30.3	
58	Buenos Aires 1906	Buenos Aires, Argentine	75—87	monthly averages

The ammonia concentration in air thus appears to be about 30 times higher than the equilibrium concentration calculated by BUCH, and the sea as a source for ammonia in air seems to be out of question, unless it is brought to the air by some other means than pure diffusion. One such possibility is the spray which is known to carry chlorides and other salts from the sea up into the air. If the amounts of different salts brought into the air by spray were in the same proportion as in the sea, however, the amount of ammonia in spray should be very low. If the spray has a different composition from the sea water an enrichment of ammonia in the spray is possible.

As a conclusion it can be stated that all evidence so far is against SCHLÖSING's theory about the sea as the great source for ammonia in air.

b. *The equilibrium in the air—soil system*

In 1872 BRETSCHNEIDER (54) published an account of experiments where he tried to estimate the amount of ammonia absorbed by the soil. These are probably the first experiments of this kind. He relates an earlier discovery by another German, SCHÖNBEIN, who found that H_4NNO_3 was formed on evaporation of water. The conclusions drawn by this discovery was that H_4NNO_3 was synthesized in the plants during their transpiration. BRETSCHNEIDER calculated, however, that a field of red clover would receive only four per cent of its nitrogen in this way. (This was also before the discovery of the nitrogen fixing ability of legumes). He found, further, that water which had been in contact with air contained an excess of ammonia over nitrate and concluded that the excess must have been absorbed from the atmosphere. He then started to investigate the absorption ability of "synthetic" soil and found that a moist mixture of 95 per cent quartz sand and 5 per cent ulmin (a humic acid prepared from sugar) absorbed 46 kg ammonia nitrogen/ha/year. Pure quartz sand absorbed only 1.6 kg and a water surface 0.8 g. His conclusion was that if SCHÖNBEIN's theory was correct the amount of nitrogen fixed in this way was of much less importance than the absorption of ammonia from the air. Very soon it was found, however, that SCHÖNBEIN's theory was incorrect, his results were due to contamination of the air in the laboratory where he made his discovery. But BRETSCHEIDER's results were really important. Similar experiments using acids instead of soil carried out in Rostock (138) in 1879—1891 gave 30.6 kg ammonia nitrogen/ /ha/year and in Tokyo (157) 11.6 kg. SCHLÖSING found, as mentioned before, that the soil absorbed 40 kg.

In 1911 HALL & MILLER (129) in Rothamsted published results of a critical study on the absorption of ammonia from air by acids. They were unable to show that soil could absorb ammonia from the atmosphere in direct experiments. By using two dishes filled with acid, mounted at different heights in the same place, they found, however, that the lower contained generally less ammonia than the upper one which they took as an indication of a steady absorption of ammonia by the soil. To this interpretation it can be objected that the difference in ammonia contents could also be due to differences in the volume of air brought in contact with the dishes. The turbulence of the air is generally increasing with the height above surface. The amount they found absorbed in the dishes were, however, quite low. The results from 1908—09 and 1909—10 for four different fields at Rothamsted are shown below, the amounts are all given in kg per hectare. The dishes were placed at 5 and 105 cm height respectively.

Field	I		II		III		IV	
	Upper	Lower	Upper	Lower	Upper	Lower	Upper	Lower
1908—1909	1.55	0.88	1.25	1.45	1.17	1.29	0.79	0.66
1909—1910	1.77	1.02	1.28	1.98	1.41	2.11	0.95	0.88

The fields II and III are manured with ammonium sulfate in the spring. This manure brought about a greater absorption of ammonia in the lower together with an 10-fold increase which persisted for some time. This explains why the "lower" figures are higher for these fields. Thus a reverse process may take place under certain conditions.

The figures in the table are, however, quite low and they drew the conclusion that the amounts of ammonia absorbed from air by the soil is of no practical importance. They had, however, the dishes covered with fine meshed brass nets to prevent dust particles from entering the dishes. This arrangement, of course, does not allow a free contact with air which they also showed by comparison with uncovered dishes. Uncovered dishes absorbed about four times as much ammonia. Their results are thus not comparable with those related before. Obviously considerable amounts of ammonia can be absorbed from the atmosphere by soils under suitable conditions.

That a reverse process may take place has also been observed by EHRENBERG (96) and SCHNEIDEWIND (266). They showed that lime-rich soils, when receiving ammonia in fertilizers, may lose ammonia to the atmosphere.

Any moist soil, with a pH lower than 7 will, however, absorb ammonia from air, the amount depending upon the ammonia content of the air, and the rate of vertical mixing, i. e. the volume of air brought into contact with unit soil surface. In this connection some recent results given by INGHAM (148) indicate that plant leaves can adsorb ammonia physically, this ammonia then being washed down to the soil by rain.

c. *The source of ammonia in air*

In their summary on the Rothamsted rainwater investigations RUSSELL & RICHARD (254) in 1919 listed the probable sources of ammonia in air as the sea, the soil and fossil deposits in which ammonia is released by combustion. They thought that the soil was probably the most important source, this conclusion being based mainly on the work of HALL & MILLER (loc. cit.) and the fact that, at least at inland places, the ammonia content in rain is higher during summer than during winter. The work by HILL & MILLER has been discussed in the preceding section and if the sea, with a pH of about 8, absorbs ammonia from the air most soils will certainly do the same. The fact that the ammonia contents are higher during summer than during winter at inland places may be due to increased convection in the air during summer, the excess over winter merely reflecting the contribution of cyclic dust during the hot season. Thus the sea and the soil cannot act as a real source for ammonia in air.

The third source, namely ammonia in fossils used as fuel is considerable. Already RUSSELL & RICHARD calculated that in Great Britain the amount of ammonia nitrogen released by combustion of coal was about 5 times higher than that actually received from rain water but they did not think it could be the only source. It is well known that the ammonia in rainwater in industrial centres is high but it generally decreases rapidly going out from the centre. Due to the general turbulence and movement of air masses the ammonia delivered from local spots will be rapidly spread out to cover extensive areas. To this source should be added the role of forest fires, as yet unexplored. As a matter of fact these sources of ammonia are the only ones possible as far as we know. Maybe in future some physico-chemical processes of ammonia formation in nature can be discovered, e. g. photo-chemical reduction of nitrate or even nitrogen to ammonia. So far, however, no processes of that kind seem to be known.

The higher European values in fig. 2 compared to non-European are most probably due to a higher coal consumption in Europe than in other countries.

To summarize, it can be stated that the only net source of ammonia in air seems to be the ammonia released on combustion of fuels. Ammonia in cyclic dust adsorbed on fine particles from the ground does not increase the ammonia pressure of the air and does not contribute anything new to the soil when washed down by rain.

7. The origin of nitrate and nitrite

The formation of nitric oxide from nitrogen and oxygen at high temperatures has been known for a long time. It was therefore quite natural to assume that nitrate in rain was derived entirely from electric discharges in the atmosphere especially as many earlier investigations reported increasing nitrate contents in rainwater during thunderstorms. Later there was, however, some doubt about this point; POSNJAKOV (239) for instance, failed to find any correlation between nitrate content and frequency of lightning; the same was also pointed out by RAM (243). Others (110, 111, 164) found that both ammonia and nitrate increased during thunderstorms. In one place with exceptionally high nitrate content, namely Hanoi, Tonkin (17) there was a positive correlation but these values are so extremely high that they may be classified as "very improbable".

HUTCHINGTON (145) has, on basis of the analyses from Oklahoma (110, 111) estimated the amount of nitrate which could be derived from electric discharges in this place and found around 14 per cent could be accounted for. The rest must be derived from some other source. This low estimate is compatible with the fact that in tropical places with a very high frequency of electric discharges the total amount of nitrate nitrogen is not exceedingly high and not at all in proportion to the frequency of lightning.

Some workers from India, namely RAO, GAPOLO & DHAR (244), DHAR & RAM (89) and RAM (243) have suggested another possible source of nitrate in the atmosphere. On examining the relative amounts of ammonia and nitrate in rainwater in temperate and tropical places it was found (see MILLER [213]) that the relative amount of nitrate nitrogen is higher in tropical places although the total nitrogen is of the same order. This they explained to be due to a photo-chemical oxidation of ammonia to nitrate in tropical places, the intense ultraviolet radiation delivering the necessary activation energy. According to RAM (243) the ammonia molecule should be oxidized by O_2 possibly through the formation of ozone, O_3 from O_2 by the action of ultraviolet radiation, the ozone being a strong oxidant.

Several objections can be raised against this theory. Firstly it is doubtful whether the nitrate fraction is higher in tropical places than in temperate ones. This has been discussed earlier in this paper (see p. 222). Secondly the amount of ultraviolet radiation received near the ground is highest at the poles, and decreases towards the equator.

There seems to be, however, a positive correlation between nitrate in rain and ozone content in air. This has at least been shown qualitively in 1876 by EUGLING (103). This could, of course, be taken as an indication that the nitric oxide formation in air is facilitated by the presense of ozone. But on the other hand the appearance of ozone at lower altitudes indicates great vertical convections in the upper atmosphere as suggested by REGENER (245). If nitrate is formed at high altitudes such great convections will bring fresh reserves of nitrate into the lower atmosphere as well, where it could easily influence the nitrate content of rainwater. So it does not seem necessary to assume that ozone itself takes an active part of the nitrate formation.

In the atmosphere there seems to be a considerable concentration of nitrous oxide. It was first discovered by ADEL (2) in 1939 from its absorption spectra in the infra red. Later it was estimated by the masspectrographic method on samples of air taken near the ground to $5 \cdot 10^{-5}$ per cent by volyme. (SLOBOD and KROGLI [276] in 1950). On the basis of these and some other results ADEL (3, 4) proposes a nitrogen cycle in which nitrous oxide takes part. He suggests that N_2O is formed in the soil by decomposition of H_4NNO_3. As N_2O is a rather inert gas it escapes into the soil air from where it diffuses into the atmosphere. When it reaches high altitudes it is decomposed by the action of ultraviolett light ($\lambda < 2{,}000$ Å) into N_2, O_2 and NO. The nitric oxide thus formed is, however, also photo-chemically decomposed

into N_2 and O_2. The nitrogen formed diffuses into the lower atmosphere where it is recombined with O_2 to NO under influence of electric discharges. The nitric oxide formed in this process is brought to the soil by precipitation where it combines with ammonia to H_4NNO_3 whereby the cycle is closed.

Several objections can be raised against this theory. The decomposition of H_4NNO_3 into N_2O and water is well known. For the pure, dry salt, however, it does not occur below 180° C. The reaction might, of course, still occur in the soils at ordinary temperatures as the result of some now unknown microbiological action, especially as the energy released during the decomposition is considerable. The presence of H_4NNO_3 in the soil is, however, also unlikely. The ammonia content in soils is generally very low, much lower than the nitrate as it is rapidly oxidized to nitrate as soon it is released on decay of organic matter.

It is more probable that the N_2O of the atmosphere is formed from nitrogen compounds in fossils in their combustion.

The decomposition of N_2O under the influence of ultraviolet radiation is known to occur, at least in presence of oxygen, when NO is formed. If the decomposition stopped at that stage the NO thus formed could well account for the nitric oxides present in the atmosphere.

That nitric oxide exists to fairly high altitudes can be inferred from an Australian investigation in 1914, ANDERSON (11, 12), MASON (202, 203) et al. They made daily analyses of nitrate and nitrite in rainwater near Melbourne and found that the content and amount of these compounds was dependent on the origin of the air masses. They grouped the values according to the weather type into the following groups

1. antarctic low pressures
2. antarctic low pressures influenced by tropical low pressures
3. tropical low pressures.

It was found that each individual rainfall of types 2 and 3 carried down an amount of nitrate and nitrite which was independent of the amount of rainfall. The tropical low pressure rains gave the greatest amount of nitric nitrogen.

Further analyses of the air for nitric nitrogen confirmed their assumption that the amount of nitric nitrogen in each rainfall was to a great extent dependent upon the vertical extension of the air masses involved. From their air analyses they calculated the height required to give an amount corresponding to a certain type of rainfall. The calculated height was of the same order of magnitude as that estimated by meteorologists. They also found that the nitric nitrogen content of air was much higher on the forward side of a low pressure system than on the rear side.

From their results it appears that nitric nitrogen should be fairly evenly distributed throughout the troposphere.

8. The geochemistry of nitrogen

Geochemical data on nitrogen gives an interesting picture of its distribution in nature. Some data are given by HUTCHINGTON (loc. cit.). The nitrogen content of igneous rocks is surprisingly constant, roughly 0.005 per cent (by weight). Most of it is present as ammonia while nitrate seems to be absent. Sedimentary rocks contain as an average 0.051 per cent, 0.046 per cent thus being biochemically fixed. The total amount present is thus

Igneous rocks	37,000	g N per cm^3
Sedimentary rocks	87	» » » »
In the atmosphere	755	» » » »

Thus it is seen that roughly 10 per cent of the nitrogen content in the atmosphere has been fixed by the biosphere during the time since life appeared. Part of this nitrogen is being released at the present time through combustion of coal but this amount is probably small. The fixation process is, of course, still going on.

9. The biological importance of combined nitrogen in air and precipitation

For most plant species a supply of combined nitrogen is essential for development, and of course, they do not care about the origin of this. The rate of delivery is the only relevant question for the plants.

It may be of some advantage to classify all the combined nitrogen brought to the soil in two groups, namely cyclic and non-cyclic,

the term cyclic taken in a short time sense. To the cyclic group belongs all nitrogen precipitated from the atmosphere as cyclic dust and nothing more. To the non-cyclic group belongs all other forms of combined nitrogen, even that released on combustion of coal.

The cyclic dust does not contribute anything new to the total land surface, on the contrary some of it may be carried out over the sea and precipitated there. This fraction represents a loss. The biological importance of cyclic dust lies in the fact that it will redistribute non-cyclic nitrogen and even out differences in the soil income of non-cyclic nitrogen.

As cyclic dust will be represented in rain-water analyses, these may well represent the total income from atmospheric precipitation but are not a reliable figure for the net income. The possibility of a direct absorption of ammonia from the air by the soil must, however, also be taken into consideration.

(To be continued in the next issue.)

Electron-Microscope Study of Cloud and Fog Nuclei

By S. OGIWARA and T. ŌKITA
Tōhoku University, Sendai, Japan

(Manuscript received 5 March 1952)

Abstract

Droplets of clouds on a mountain and of fog in an urban area were captured and the form, nature and size of their nuclei were studied by means of an electron-microscope and by a chamber of constant humidity. These nuclei have similar form and nature to the hygroscopic particles in haze and to the artificially produced combustion particles. No sea-salt nuclei were found in our observations, therefore sea-spray appears to be an insignificant source of condensation nuclei. It was found that both the cloud and the fog nuclei originated in combustion products which were the mixture of hygroscopic and non-hygroscopic substances, and that the greater part of the nuclei did not contain pure sulfuric acid.

I. Introduction

The nature and origin of cloud and fog nuclei have been studied by many investigators and SIMPSON (1941) concluded that "sea-salt, sulfuric acid and nitrous acid drops were important nuclei in the atmosphere". However, these nuclei are so small that their size and nature cannot be studied by an optical microscope. Recently KUROIWA (1951) observed the sea fog nuclei by an electron microscope and found that many of them were solid particles insoluble in water and that only a small fraction of them were sea-salt particles. No sea-salt nuclei were found on Mt. Niseko (1,300 m) by the same author (KUROIWA and TADANO 1951). However, he did not refer to the origin of the nuclei.

From May to July, 1951 we observed the cloud nuclei on Mt. Zaō and the fog nuclei in Sendai (fig. 1). Observations and experiments were made mainly by an electron microscope and by a chamber of constant humidity. To study the nature and origin of nuclei we compared them with haze particles, sea-salt particles and artificially produced combustion particles.

II. Observation of cloud and fog nuclei

1. *Capture of cloud and fog droplets*

Cloud and droplets were captured on a thin collodion film coated on the sample holder (Träger) of the electron microscope. As the droplets thus captured on the Träger evaporated quite rapidly the following technique was employed:

A small metal case was placed on the stage of an optical microscope as shown in fig. 2 (a) and the Träger was put into it. Drawing the air by a syringe which was connected to the case, the droplets in the air were captured upon the collodion film and their locations on the film were easily detected by the microscope.

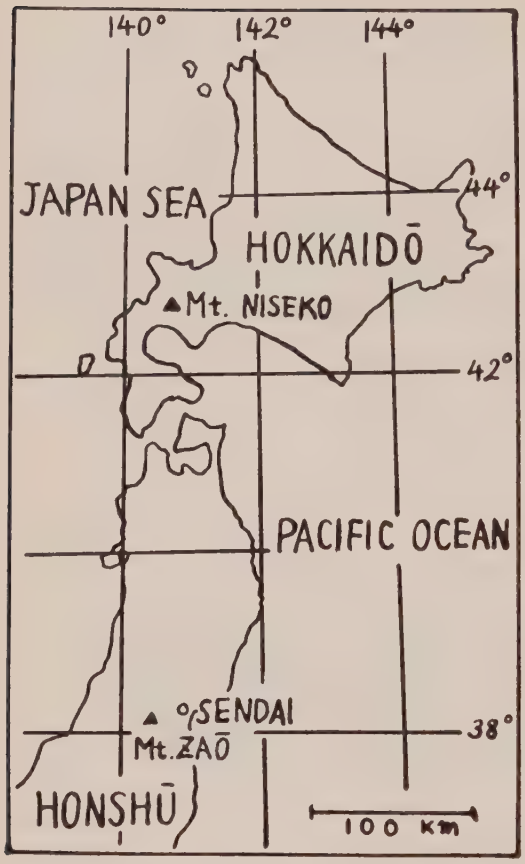


Fig. 1.

The total number of droplets captured were 60 on Mt. Zaō and 15 in Sendai, respectively. The time and station of observation and the weather conditions are shown in Table 1.

2. Nature of nuclei.

The size of the nuclei which remained on the collodion film after the cloud and fog droplets evaporated was measured by the optical microscope. The humidity in the room was 60 to 70 % and the magnification of the microscope was 900. Some of the nuclei were too small to be detected by the optical microscope.

Then the nuclei were put into a small chamber of constant humidity shown in fig. 2 (b). The chamber was a tapered glass vessel whose diameter was about 4 cm, and on its bottom a sheet of filter paper soaked with distilled water was placed. In the chamber almost all nuclei increased their diameter by absorbing the water vapour and the increment continued for 3 or 4 minutes. The final diameter was about 2 to 3 times as large as the initial one. On the small nuclei, which could not be detected at the room humidity, water droplets were also formed. Two nuclei, however, retained their original size in the chamber.

The temperature of the filter paper was equal to room temperature and the experiment was made in an underground dark room to avoid temperature fluctuations. Therefore the humidity in the chamber will be considered as just 100 %. As the nuclei which grew in the chamber are able to form cloud and fog droplets in the atmosphere with a humidity

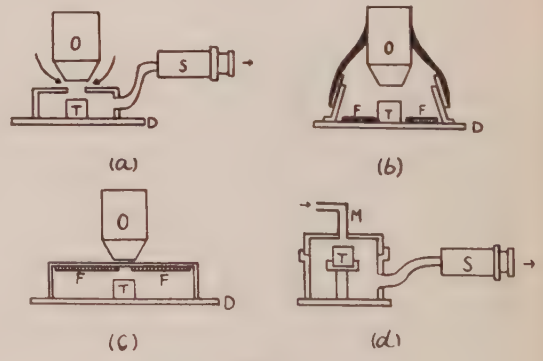


Fig 2.

Table 1. Observation of cloud and fog.

Date	Time h m	Station of Observation	T_{dry} (°C)	T_{wet} (°C)	R. H. (%)	Visibility (m)	Wind Direction	Cloud Form
11 May	03 30	Tōhoku University.....	8.8	8.8	100	—	—	Fog
5 June	07 00	» »	17.4	17.2	98	250	—	»
15 »	04 00	» »	17.5	17.3	98	250	—	»
17 »	05 00	» »	15.4	15.4	100	250	—	»
19 July	05 50	Mt. Zaō (Hütte 1,400 m) .	13.0	13.0	100	200	WSW	Cu
23 »	08 00	Mt. Zaō (Mt. Katta 1,750 m	—	—	—	—	»	»

of 100 % or below we shall refer to them as "hygroscopic nuclei". The two nuclei referred to above are incapable of acting as condensation nuclei except in the supersaturated atmosphere, shall be referred to as "non-hygroscopic nuclei". The above results show that cloud and fog nuclei are formed in the atmosphere with a humidity of 100 % or below. There was one cloud sample in which no nuclei were found either by the optical or by electron microscope.

After the above experiment was finished, the form and size of the nuclei were observed by the electron microscope. Miscellaneous forms of the nuclei were observed and typical forms of them are shown in the photographs on page 236 and 237. Those from (1) to (13) are hygroscopic nuclei and (14) is a non-hygroscopic one.

(1) . . . Nearly circular thin filmy substance, in about half of this type of nuclei 1 to 10 small dark spots are also seen. These nuclei range in diameter from 0.3 to 1 μ.
(3) and (6) . . . A large well defined dark particle and thin filmy substance around it. The former is of polygonal, circular or ragged form and its diameter is 0.3 to 1.2 μ.
(8) . . . Special form as seen in the photograph.
(9) . . . Spotted thin substances, in the centre of which a dark particle is seen.
(10) and (12) . . . Resemble (3) etc. . . As will be described later, however, they do not readily lose their hygroscopic nature by electron-microphotographing.
(14) . . . Non-hygroscopic nucleus. Presumably it is a clump of carbon particles of oil as shown in (15). It should be noted that most nuclei belong to the former two types.

After the electron microphotographs of the specimens were taken, the hygroscopic nuclei were again put into the chamber of constant humidity. Then it was found unexpectedly that the moisture in the chamber did not condense on the greater part of the nuclei, that is to say, these nuclei had then lost their hygroscopic nature. It is supposed that by electron bombardment the hygroscopic substances contained in the nuclei were evaporated during photographing. Examples of these nuclei are shown in photos (1), (3), (6), (8) and (9). On the other hand three nuclei did not readily lose their hygroscopic nature, two of which are shown in (10) and (12). It should be noted that whether the nuclei lost their hygroscopic nature or not did not primarily depend upon their size, since, even some small nuclei did not necessarily lose their hygroscopic nature, while some nuclei with large diameters lost it by photographing. These results may be explained by assuming that generally several kinds of hygroscopic substances are contained in the nuclei and the rate of evaporation of these substances by electron bombardment differs from one kind to another.

To see whether the residues of the nuclei which lost their hygroscopic nature were soluble in water the following experiment was made. The residues were placed in a sealed glass vessel lined with moistened filter-paper as shown in fig. 2 (c). When the temperature of the filter paper was higher than the room temperature the moisture in the vessel condensed on the residues and water droplets were seen by the optical microscope. After the residues were wetted with water like this and then dried, their photographs were again taken by the electronmicroscope.

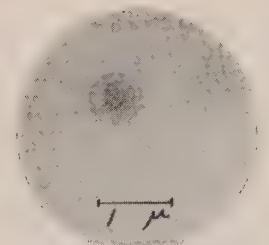
Comparing the image of the nuclei before and after wetting, it was found that most nuclei retained their original form as shown in (1) and (2), (3) and (4), (6) and (7). In three nuclei which retained their hygroscopic nature, however, the filmy substance changed form considerably, as shown in (10) and (11), (12) and (13).

From the experimental study described above it may be concluded that cloud and fog nuclei are composed of filmy substances and non-hygroscopic solid particles. A part of the substances with thin filmy appearance are also non-hygroscopic and insoluble in water, while others are hygroscopic and soluble in water.

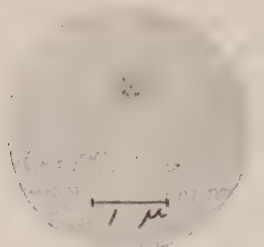
III. Observation of haze particles

1. Capture of particles

For comparison with cloud and fog nuclei, haze particles were captured in Sendai. Near the ground there were large particles exceeding several microns in diameter. However, most particles captured on the roof, about 10 m above the ground, were comparable in



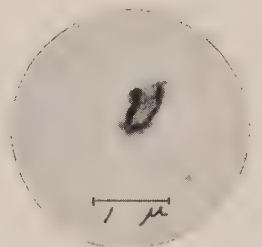
(1) *Hygroscopic nucleus, Mt. Zaö.*



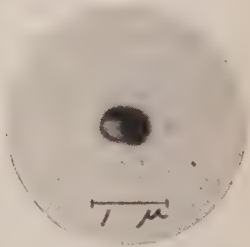
(2) *Nucleus (1) after rewetting.*

(3) *Hygroscopic nucleus, Mt. Zaö.*

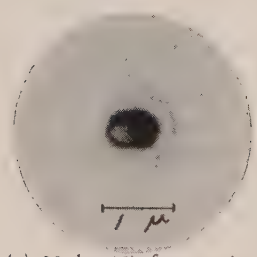
(4) *Nucleus (3) after rewetting.*



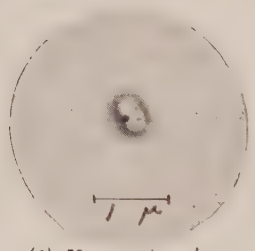
(5) *Nucleus (4) after reaction with HF.*



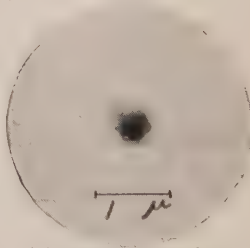
(6) *Hygroscopic nucleus, Sendai.*



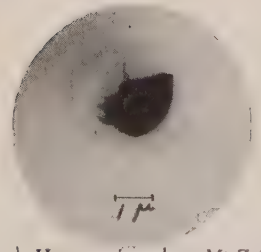
(7) *Nucleus (6) after rewetting.*



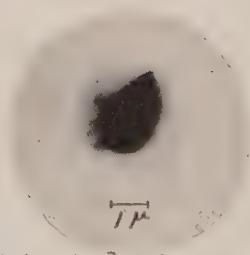
(8) *Hygroscopic nucleus, Mt. Zaö.*



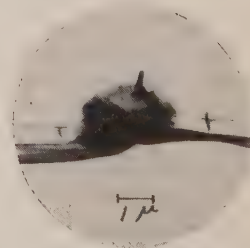
(9) *Hygroscopic nucleus, Sendai*



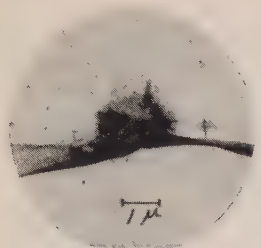
(10) *Hygroscopic nucleus, Mt. Zaö.*



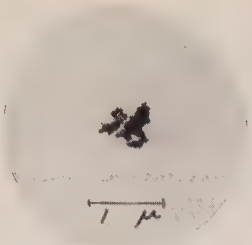
(11) *Nucleus (10) after rewetting*



(12) *Hygroscopic nucleus, Mt. Zaö.*



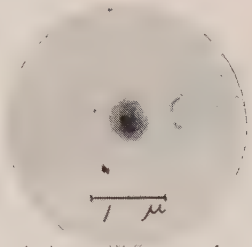
(13) *Nucleus (12) after rewetting*



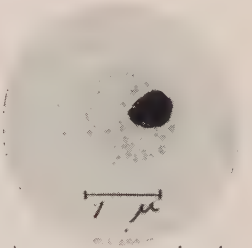
(14) *Non-hygroscopic nucleus, Sendai.*

(15) *Smoke particles of light oil.*

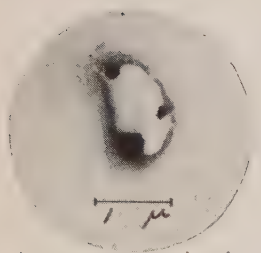
(16) *Non-hygroscopic particles in haze.*



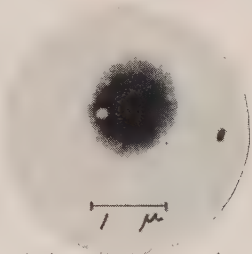
(17) *Hygroscopic particle in haze.*



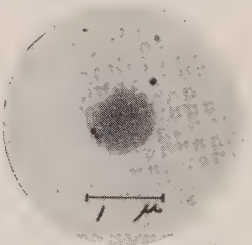
(18) *Hygroscopic particle in haze*



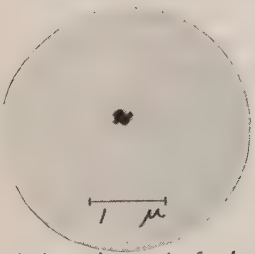
(19) *Hygroscopic particle in haze.*



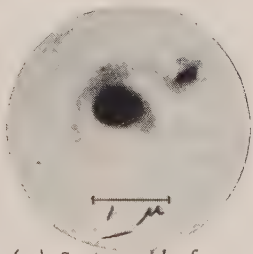
(20) *Hygroscopic particle in haze*



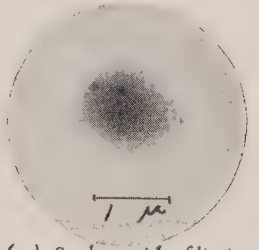
(21) *Smoke particle of Japanese cypress.*



(22) *Smoke particle of coal*



(23) *Smoke particle of grass.*



(24) *Smoke particle of lignite.*

Table 2. Observation of haze (1951).

Date	Time	Number of Particles		T_{dry}	T_{wet}	R. H.	Visibility	Wind	Cloud	
	h m	Hygroscopic	Non-hygroscopic	(°C)	(°C)	(%)	(m)		Amount	Form
19 June	05 30	14	13	—	—	—	—	—	—	—
19 »	06 00	11	40	18.0	14.3	67	—	Calm	9	as, ci, cc
20 »	06 30	3	14	19.8	16.2	69	—	»	8	ac, as, sc, ci
23 »	05 40	9	16	17.5	14.4	72	2,500	»	1	cs, ci, cu
24 »	05 40	14	18	15.5	14.0	85	3,500	—	0	—
25 »	06 00	35	35	17.4	15.3	80	2,000	NW	5	cs, ci
4 Sep.	06 40	41	154	19.3	16.9	79	4,000	WNW	8	ac, as, cu
9 »	07 00	10	37	20.9	18.7	81	7,000	—	10	sc
23 Oct.	06 30	1	47	6.3	5.2	84	1,200	NNW	9	cs, ac, sc

size with the cloud and fog nuclei. To capture these small particles we used a collector shown in fig. 2 (d), in which the particles passed through a metal tube *M* at high speed and struck the collodion film, to which they then adhered.

The time and the weather conditions are shown in Table 2 together with the number of haze particles captured by the collector. It will be seen from the table that the haze is composed of hygroscopic and non-hygroscopic particles in various proportions. Experimental procedure to decide whether the particles were hygroscopic was quite the same as that for cloud and fog nuclei.

2. *Nature of haze particles*

Electron-microphotographs of hygroscopic haze particles are shown in (17) to (20). As will be seen by comparing (17) with (1), (18) with (6), and (19) with (8) they are similar in form to the hygroscopic nuclei of clouds and fog. Furthermore these particles lost their hygroscopic nature by photographing. Thus hygroscopic particles in haze are similar both in form and in nature to cloud and fog nuclei.

Two hygroscopic particles evaporated completely as a result of the electron bombardment and left nothing on the collodion film. These particles probably consisted entirely of hygroscopic substances.

One example of non-hygroscopic particles in haze is shown in photo (16). There are three particles one of which is similar in form to the smoke particles of fuel (?) oil and others have circular form. Since these particles remained unchanged under exposure to hydrogen fluoride vapour it is presumed that they consisted of carbon.

IV. Origin of cloud and fog nuclei

1. *Comparison with sea-salt particles*

Is has been considered that a large number of sea-salt particles are supplied from the sea-spray and that they are the most important condensation nuclei in the atmosphere. To decide whether the cloud and fog nuclei captured by us were sea-salt particles, small sea-salt particles with a diameter of about 0.5 μ were produced by spouting a water solution of sea-salt. Following the same experimental procedure as described before it was found that these salt particles had images of comparatively well defined crystalline form. Furthermore it was found that they were not only soluble in water but also did not lose their hygroscopic nature by electron-microphotographing. Thus the sea-salt particles have different form and nature from the greater part of cloud and fog nuclei.

As already described, however, a few nuclei of cloud and fog did not lose their hygroscopic nature by electron bombardment. But hygroscopic substances contained in these nuclei were composed of those with thin filmy appearance, which were far from salt crystals.

Recently KUROIWA and TADANO (KUROIWA 1951, KUROIWA and TADANO 1951) observed the nuclei of clouds on Mt. Niseko and those of sea fog on the coast of Hokkaido by using an electron microscope. They found no sea-salt nuclei on Mt. Niseko, and even in the sea fog only a small fraction of the nuclei were sea-salt particles. These results lead one

to the conclusion that sea-spray is not an important source of condensation nuclei of cloud and fog, at least not in these cases.

It is well known, however, that a considerable amount of sodium chloride is contained in cloud and fog water (KÖHLER 1922, HOUGHTON and RADFORD 1938, MIYAKE 1948), and the sea-salt nucleus theory has been based on this fact. Therefore it is necessary to explain the origin of sodium chloride in cloud and fog water. From some preliminary experiments sea-salt particles were found to have equal hygroscopic activity to combustion particles of the same size. Observations of sea-salt particles on the coast of Honshū (OGIWARA and ŌKITA 1951) showed that the number of particles larger than cloud and fog nuclei was about 5 per c.c. and was far less than that of cloud and fog droplets in the atmosphere. Therefore it seems reasonable to conclude that the greater part of cloud and fog droplets are formed on combustion nuclei, and that small sea-salt particles cannot act as condensation nuclei in the atmosphere. When cloud or fog water is collected by wire screen or other kinds of collectors, small sea-salt particles floating in the atmosphere should be captured together with the cloud or fog droplets. Large particles, which would become condensation nuclei, but are not detected in the cloud and fog droplets owing to the smallness of their number, would be added to the sodium chloride content in the cloud and fog water. Quantitative explanation for the sodium chloride content requires further investigation.

2. *Comparison with combustion products*

As it was found that cloud and fog nuclei are not sea-salt particles but are similar in form and nature to hygroscopic haze particles, the most probable source of the nuclei is considered to be combustion. Smoke particles produced by combustion of many kinds of fuel, such as coal, lignite, leaf of Japanese cypress, pine tree and grass were observed following the method described in chapter II. Some of these smoke particles are shown in photos (21) to (24). It was found that they too are a mixture of hygroscopic and non-hygroscopic substances and that they may lose their hygroscopic nature by electron-microphotographing. Cloud and fog nuclei are therefore similar in form and nature to combustion products. From the results described it may be concluded that most nuclei of cloud and fog originate in combustion.

It remains a question, whether the non-hygroscopic particles contained in cloud or fog nuclei are those captured by the cloud or fog droplets which have formed on the hygroscopic nuclei. As cloud and fog nuclei have similar form, nature and size to combustion products, the greater part of these non-hygroscopic particles are considered to have been adhering to the hygroscopic substances from the beginning.

3. *Chemical composition of combustion nuclei*

Generally combustion products are composed of many kinds of substances, therefore it is not easy to study their chemical composition in detail.

It has been considered that sulfur dioxide produced by combustion of coal etc. becomes sulfuric acid droplets by oxidation, and that these hygroscopic droplets can act as effective condensation nuclei of cloud and fog. To decide whether the hygroscopic substance in the nuclei is sulfuric acid the following experiments were made.

When a droplet of sulfuric acid with a diameter of about 0.5 μ was put into the chamber of constant humidity it increased its diameter by 2 to 3 times. Thus the increment of the droplet was of quite the same order as that of cloud and fog nuclei. Therefore the quantity of hygroscopic substances in cloud and fog nuclei cannot be considered to be far smaller than that of sulfuric acid. The droplet was then exposed to the electron beam of the microscope. Even after the electron bombardment, it was found that the droplet did not change its size nor its hygroscopic nature. From the above results it is obvious that the greater part of cloud and fog nuclei do not contain sulfuric acid. Chemical engineers have found that many kinds of sulfuric compound are contained in the smoke of coal and that some of them are hygroscopic. Therefore it may be considered that the sulfuric ions dissolved in cloud and fog water would originate in these sulfuric compounds but not in pure sulfuric acid droplets.

Next, to study the chemical composition of the nonhygroscopic solid particles in the nuclei, they were exposed to the vapour of hydrogen fluoride. Some particles were attacked by the vapour as shown in photo (5), while others were unchanged. By the same technique it was found that the solid particles in the smoke of light oil, leaf of Japanese cypress and grass were proof against hydrogen fluoride, while some of those from smoke of coal and lignite were attacked by hydrogen fluoride. It is presumed that the former is a pure carbon particle and the latter is a silicate or carbonate, and that the solid particles in the nuclei are composed of these substances.

V. Conclusion

From the investigations set forth above it is concluded that the greater part of cloud and fog nuclei are composed of hygroscopic and non-hygroscopic substances which originate in combustion and that sea-spray is not a significant source of condensation nuclei. It was also shown that the greater part of the hygroscopic substances contained in the nuclei are not pure sulfuric acid. It is important to bear in mind that, contrary to the opinion which has been vigorously advocated so far, the greater part of cloud and fog nuclei are composed of neither sea-salt nor sulfuric acid.

A question will arise: If the combustion products are the most important nuclei, what was the most effective nucleus before man used fire? Even at that time combustion nuclei would be supplied by the eruption of volcanos or by forest fires. Or it may be more probable to consider that cloud and fog droplets would be formed on other kinds of less effective nuclei such as small sea-salt nuclei at a slightly higher atmospheric humidity than at present. As it has been assumed that by the appearance of life the chemical composition of the air has changed its proportions, it is perhaps no too bold to assume that by the appearance o man the most effective nucleus had changed from sea-salt to combustion products.

The authors thank the Department of Education for financial support and to Prof. T. Hibi and Mr. S. Takahashi of the Research Institute for Scientific Measurement of the Tōhoku University for their assistance with electron-microphotographing. Thanks are also due to Dr. B. Tadano of the Hitachi Central Research Laboratory for loaning us the electron-microphotographs of smoke.

REFERENCES

Houghton, H. G., and Radford, W. H., 1938: On the Measurement of Drop Size and Liquid Water Content in Fogs and Clouds. *Papers in Physical Oceanography and Meteorology*, **6**, No. 4.

Köhler, H., 1922: Zur Kondensation des Wasserdampfes in der Atmosphäre. 1 and 2. *Geofysiske Publikationer*, **2**, 1 and 2.

Kuroiwa, D., 1951: Electron-Microscope Study of Fog Nuclei. *Journ. Meteor.*, **8**, 157—160.

Kuroiwa, D., and Tadano, B., 1951: Investigations on the Condensation Nuclei of Fog by an Electron-Microscope. *Low Temperature Science*, **7**, 51—66.

Miyake, Y., 1948: The Chemical Nature of the Saline Matter in the Atmosphere. *Geophysical Magazine*, **16**, 64—65.

Ogiwara, S., and Ōkita, T., 1951: The Size Distribution of Sea-Salt Nuclei and Its Contribution to Visibility. *Science Report Tōhoku Univ.*, Series **5**, 3, 78—86.

Simpson, G. C., 1941: On the Formation of Cloud and Rain. *Quart. Journ. Roy. Met. Soc.*, **67**, 99—133.

Sunspot Areas, Flares and Filaments observed at the Stockholm Observatory in the Years 1950 and 1951

By Y. ÖHMAN and L. O. LODÉN
Stockholm Observatory

(Manuscript received 15 June 1952)

Abstract

Tables are presented giving sunspot areas, flares and filaments according to the solar routine observations carried out at the Stockholm Observatory in 1950 and 1951. The material has been distributed before as current reports to a number of Swedish institutions. A brief discussion is given and a description is made of the methods used for the announcements of the general solar activity.

A. *Sunspots and faculae.* In an earlier publication[1] a description has been given of the routine solar work started at the Stockholm Observatory in 1947 in order to furnish Swedish geophysical institutions and the Royal Swedish Board of Telegraph with solar data in addition to the information given in the Ursigrams. The photographic observations of sunspots have been made by us in the same way as before, that is by using a Zeiss objective with an aperture of only 30 mm and a focal length of 800 mm and by enlarging the solar image to a diameter of 56 mm. A passband of 200 Å is selected in the H and K region of the spectrum by means of a Schott and Gen. U G 2 filter with a thickness of 1 mm. The photographs are made on Ilford Lantern contrasting plates.

The solar photographs obtained in this way have a remarkably uniform quality as the small aperture reduces the influence of variations in seeing-conditions. Moreover the ultraviolet filter brings out a high contrast. The spots are very dark and outstanding objects and the bright faculae can be traced not only close to the limb but all over the surface of the sun. In fact, the accumulated line absorption in the violet seems to produce the characteristics of the low chromosphere though with less contrast than on real spectroheliograms obtained in the metallic lines and in the outer wings of the H and K lines.

In table I and II a presentation is given of sunspot areas for 1950 and 1951 according to our measurements. These tables are arranged in the same way as those for 1948 and 1949 given in the previous publication[1]. The sunspot areas are thus measured in 10^{-5} of the half sphere of the sun as unit and the figures refer to the first day-sector after the sun's central meridian observed at about $8^h\ 30^m$ U. T. The upper number in the tables refers to the north hemisphere and the lower number to the south hemisphere. When a number is given within parenthesis no ob-

[1] Tellus, **2**, 255, 1950. (Stockholms Observatorium Meddelande No. 65).

[1] Loc. cit.

Table I. 1950.

Sunspot areas in the day-sector after the Sun's central meridian together with signs of active groups + or non active groups —. Interpolated values are given within parenthesis.

	Jan.	Febr.	March	April	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.
1	—	(0) (+)	2 —	7 (—)	0 —	8 (—)	0 —	0 —	0 (—)	(0) (—)	—	(0) (—)
	—	(0) (—)	0 —	0 (—)	23 +	0 (—)	0 —	0 —	1 (—)	(0) (—)	—	(20) (—)
2	—	(0) (—)	0 —	(0) (—)	10 —	50 (—)	0 —	0 —	0 —	0 —	—	0 (—)
	—	(0) (—)	5 +	(21) (+)	2 —	0 (—)	0 —	0 —	0 —	0 —	—	5 (—)
3	—	0 (—)	(0) (—)	0 (—)	20 +	0 (—)	0 —	0 —	12 +	0 —	—	(25) (—)
	—	0 (—)	(0) (—)	0 (—)	0 (—)	0 (—)	0 —	0 —	0 —	0 —	—	(2) (—)
4	0 (—)	(0) (—)	12 —	77 +	20 (+)	0 (—)	0 —	18 —	0 —	0 (—)	8 —	0 —
	0 (—)	(0) (—)	0 —	32 +	0 (—)	0 (—)	0 —	10 +	0 —	0 (—)	0 —	0 —
5	0 —	(2) (—)	0 —	(0) (—)	7 —	0 (—)	0 —	0 —	0 —	0 —	0 —	0 —
	0 —	(0) (—)	0 —	(0) (—)	0 —	0 (—)	0 —	70 +	0 —	10 —	0 —	0 —
6	0 —	(0) (—)	0 —	8 (—)	0 —	0 —	41 +	0 —	0 (—)	19 —	(0) (—)	(0) (—)
	0 —	(0) (—)	40 +	0 (—)	0 —	0 —	0 —	0 —	0 (—)	0 —	(0) (—)	(0) (—)
7	(0) (—)	(0) (—)	0 —	6 +	14 +	0 (—)	0 —	0 —	0 —	0 —	—	0 —
	(0) (—)	(0) (—)	16 +	0 —	0 —	0 (—)	0 —	20 +	0 —	0 —	—	0 —
8	(0) (—)	0 —	27 +	0 —	0 —	0 —	0 —	0 —	0 —	0 (—)	—	0 —
	(0) (—)	0 —	0 —	0 —	2 +	0 —	0 —	0 —	0 —	0 (—)	—	5 —
9	(4) (—)	(0) (—)	0 —	(0) (—)	1 —	0 —	4 +	0 —	(0) (—)	19 —	—	0 —
	(0) (—)	(0) (—)	0 —	(0) (—)	0 —	0 —	0 —	2 —	(0) (—)	0 —	—	0 —
10	22 (+)	0 (—)	0 —	0 —	0 —	0 (—)	0 —	0 —	0 —	1 —	—	(32) (+)
	0 (—)	0 (—)	0 —	25 +	1 —	0 (—)	0 —	0 —	0 —	0 —	—	(10) (+)
11	(0) (—)	(0) (—)	(0) (—)	(0) (—)	0 +	0 (—)	(0) (—)	0 —	0 —	0 —	(0) (—)	—
	(0) (—)	(4) (—)	(1) (—)	(0) (—)	1 —	0 (—)	(0) (—)	0 —	0 —	0 —	(0) (—)	—
12	(0) (—)	(0) (—)	57 (+)	0 —	0 —	0 —	(0) (—)	21 +	(0) (—)	0 —	0 —	—
	(0) (—)	(0) (—)	0 (—)	0 —	0 —	0 —	(17) (+)	6 —	(11) (+)	0 —	0 —	—
13	0 —	(0) (—)	6 —	70 (+)	45 +	0 —	0 —	0 —	0 (—)	0 —	(0) (—)	—
	0 —	(0) (—)	55 —	0 (—)	0 —	0 —	0 —	0 —	6 (—)	0 —	(0) (—)	—
14	(5) (—)	(16) (—)	0 —	(78) (+)	0 (—)	0 (—)	1 +	0 —	0 —	0 (—)	0 —	—
	(0) (—)	(34) (+)	27 +	(0) (—)	0 (—)	0 (—)	0 +	0 —	0 —	0 (—)	0 —	—

15	(0) (—) (0) (—)	0 — 25 +	0 — 0 —	0 — 0 —	0 (—) 0 (—)	0 — 83 +	0 — 0 —	0 — 0 —	0 (—) 0 (—)	(0) (—) (24) (+)	25 + 0 —	— —
16	(0) (—) (0) (—)	(0) (—) (0) (—)	0 — 1 —	22 + 0 —	0 — 0 —	0 — 0 —	13 + 2 +	32 + 0 —	0 (—) 0 (—)	(0) (—) (0) (—)	(36) (+) (0) (—)	— —
17	0 (—) 0 (—)	(28) (+) (17) (—)	(13) (+) (0) (—)	(0) (—) (0) (—)	0 — 0 —	0 — 0 —	0 — 10 +	20 + 0 —	0 — 0 —	(0) (—) (0) (—)	(0) (—) (0) (—)	— —
18	5 (—) 0 (—)	(0) (—) (0) (—)	25 (+) 0 (—)	(0) (—) (0) (—)	0 — 0 —	0 — 5 +	0 — 0 —	6 — 0 —	20 + 11 +	0 — 0 —	(0) (—) (0) (—)	0 — 0 —
19	0 (—) 0 (—)	38 (+) 0 (—)	18 + 0 —	45 (+) 0 (—)	0 — 0 —	0 (—) 2 (—)	0 — 0 —	(0) (—) (0) (—)	5 + 0 +	0 — 0 —	0 — 0 —	— —
20	(30) (—) (0) (—)	150 (+) 0 (—)	20 (—) 0 (—)	24 + 0 —	0 — 0 —	2 — 0 —	50 + 0 —	0 — 0 —	0 (—) 0 (—)	0 — 0 —	(0) (—) (0) (—)	— —
21	(4) (—) (22) (—)	55 + 0 —	0 (—) 0 (—)	0 — 0 —	(0) (—) (6) (+)	0 — 0 —	15 — 0 —	0 — 40 +	25 + 0 —	0 (—) 0 (—)	(0) (—) (0) (—)	— —
22	(4) (—) (8) (—)	0 — 0 —	0 — 0 —	0 — 0 —	0 — 3 —	30 (+) 0 (—)	31 + 0 —	34 + 0 —	0 — 0 —	0 — 1 —	(0) (0)	— —
23	(27) (+) (0) (—)	0 — 0 —	0 — 0 —	0 — 0 —	(30) (+) (7) (—)	0 — 0 —	(0) (—) (0) (—)	0 — 0 —	0 — 0 —	0 — 0 —	(0) (0)	— —
24	(0) (—) (0) (—)	12 + 0 —	(0) (—) (12) (+)	(0) (—) (0) (—)	0 — 33 +	3 (—) 0 (—)	0 (—) 0 (—)	0 — 0 —	0 (—) 0 (—)	0 — 0 —	(0) (0)	— —
25	29 + 0 —	12 + 0 —	(12) (—) (0) (—)	0 (—) 0 (—)	0 — 0 —	39 (+) 0 (—)	0 — 60 +	(0) (—) (15) (+)	0 — 0 —	0 — 0 —	0 0	— —
26	(0) (—) (1) (—)	20 + 0 —	(0) (—) (0) (+)	(0) (—) (78) (+)	59 (+) 0 (—)	21 — 0 —	9 — 2 +	0 + 2 —	(0) (—) (0) (—)	0 — 0 —	(0) (1)	0 — 0 —
27	47 + 2 +	0 — 5 +	3 + 0 —	(0) (—) (68) (+)	29 — 1 —	2 + 10 +	(0) (—) (4) (+)	(0) (—) (0) (—)	(5) (—) (0) (—)	1 + 0 —	0 0	0 — 0 —
28	(7) (+) (0) (—)	0 — 0 —	8 + 0 —	0 (—) 39 (+)	0 (—) 0 (—)	(0) (—) (0) (—)	0 — 0 —	0 — 0 —	(0) (—) (0) (—)	(0) (—) (0) (—)	(6) (0)	30 + 0 —
29	10 + 0 —		0 — 0 +	0 (—) (12) (—)	0 (—) 6 (—)	0 — 1 +	0 — 0 —	0 — 0 —	0 — 0 —	— —	(0) (0)	0 — 0 —
30	(0) (—) (0) (—)		0 + 0 —	0 — 8 —	0 (—) 0 (—)	(0) (—) (0) (—)	(0) (—) (10) (+)	0 — 0 —	0 — 0 —	— —	0 3	0 — 0 —
31	0 — 0 —		0 — 0 —		0 (—) 0 (—)		0 — 5 +	0 — 4 +		— —		(0) (—) (0) (—)

Table II. 1951.

Sunspot areas in the day-sector after the Sun's central meridian together with signs of active groups + or non active groups —. Interpolated values are given within parenthesis

	Jan.	Febr.	March	April	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.
1	0 —	3 —	(0) (—)	(0) (—)	(0) (—)	0 —	(0) (—)	0 —	0 —	10 +	(0) (—)	(2) (+)
	0 —	0 —	(0) (—)	(0) (—)	(2) (—)	0 —	(0) (—)	0 —	0 —	0 —	(30) (+)	(0) (—)
2	(0) (—)	(0) (—)	0 —	(0) (—)	(0) (—)	0 —	0 —	0 (—)	0 —	0 —	8 +	0 —
	(0) (—)	(0) (—)	0 —	(0) (—)	(2) (—)	0 —	0 —	4 (+)	0 —	0 —	4 +	0 —
3	—	(0) (—)	0 —	(0) (—)	0 —	0 —	0 —	0 —	0 —	0 —	0 —	0 —
	—	(0) (—)	2 +	(0) (—)	0 —	0 —	0 —	0 —	0 —	0 —	0 —	0 —
4	—	0 —	(1) (+)	(0) (—)	0 +	0 (—)	0 —	0 —	0 —	0 —	(0) (—)	(0) (—)
	—	3 —	(1) (+)	(0) (—)	0 —	0 (—)	0 —	0 —	0 —	0 —	(0) (—)	(0) (—)
5	—	—	0 —	0 —	0 —	0 —	0 —	(0) (—)	(0) (—)	0 —	—	(0) (—)
	—	—	0 —	3 —	0 —	0 —	0 —	(0) (—)	(0) (—)	0 —	—	(0) (—)
6	—	—	(0) (—)	(10) (+)	(0) (—)	(0) (—)	0 —	0 —	11 +	0 —	—	0 —
	—	—	(0) (—)	(0) (—)	(0) (—)	(0) (—)	0 —	0 —	3 +	0 —	—	0 —
7	—	—	0 —	8 +	0 —	0 —	0 —	7 +	0 —	(0) (—)	—	0 —
	—	—	0 —	0 —	0 —	28 +	0 —	0 —	0 —	(17) (—)	—	0 —
8	(0) (—)	—	(0) (—)	0 (—)	0 —	0 —	(0) (—)	0 —	0 —	0 —	—	(0) (—)
	(0) (—)	—	(1) (—)	0 (—)	0 —	26 +	(0) (—)	0 —	0 —	0 —	—	(0) (—)
9	0 —	—	0 —	3 —	0 —	0 —	(5) (+)	0 —	0 (—)	0 —	—	(0) (—)
	0 —	—	0 —	0 —	0 —	0 —	(0) (—)	0 —	0 (—)	0 —	—	(0) (—)
10	0 —	—	0 —	(0) (—)	0 —	0 —	1 +	12 (+)	0 —	0 —	—	6 (—)
	0 —	—	0 —	(0) (—)	0 —	0 —	0 —	20 (+)	40 +	0 —	—	0 (—)
11	(0) (—)	—	0 —	(0) (—)	0 —	(15) (+)	0 —	(29) (+)	0 —	0 —	—	1 —
	(0) (—)	—	0 —	(10) (+)	0 —	(0) (—)	0 —	(0) (—)	0 —	0 —	—	0 —
12	—	0 —	11 (—)	0 —	0 —	85 +	0 —	0 (—)	1 —	(0) (—)	—	(0) (—)
	—	0 —	0 (—)	0 —	3 —	0 —	0 —	0 (—)	0 —	(0) (—)	—	(0) (—)
13	—	(0) (—)	(0) (—)	(0) (—)	(0) (—)	9 +	3 +	0 —	0 (—)	(0) (—)	—	(0) (—)
	—	(0) (—)	(0) (—)	(1) (—)	(0) (—)	0 —	0 —	0 —	0 (—)	(0) (—)	—	(0) (—)
14	—	(4) (—)	(0) (—)	0 —	14 +	32 (+)	0 (—)	0 —	3 (—)	(2) (+)	(0) (—)	(0) (—)
	—	(0) (—)	(10) (—)	0 —	0 —	0 (—)	54 (+)	6 +	0 (—)	(54) (+)	(0) (—)	(0) (—)

15	— —	0 (—) 0 (—)	(0) (—) (0) (—)	0 (—) 0 (—)	0 (—) 0 (—)	0 — 0 —	(0) (—) (25) (+)	0 — 0 —	33 + 0 —	0 — 2 +	0 — 0 —	(0) (—) (0) (—)
16	— —	(0) (—) (2) (—)	0 — 0 —	0 — 0 —	140 + 0 —	0 — 0 —	0 — 0 —	0 — 0 —	0 — 3 +	0 — 0 —	(0) (—) (0) (—)	0 — 0 —
17	— —	0 (—) 0 (—)	0 — 0 —	0 — 18 +	102 + 0 —	0 — 0 —	0 — 0 —	0 — 0 —	0 — 0 —	14 + 0 —	— —	(0) (—) (0) (—)
18	— —	(12) (—) (0) (—)	0 — 0 —	54 + 0 —	0 — 0 —	(0) (—) (0) (—)	0 — 0 —	0 — 0 —	1 — 0 —	0 — 0 —	— —	(0) (—) (0) (—)
19	— —	0 — 0 —	(0) (—) (2) (—)	65 + 2 —	5 + 0 —	7 + 76 +	0 — 0 —	0 — 8 +	(0) (—) (0) (—)	11 + 0 —	(0) (—) (0) (—)	(3) (+) (0) (—)
20	(0) (—) (0) (—)	0 (—) 0 (—)	(0) (—) (0) (—)	0 — 0 —	(0) (—) (0) (—)	0 — 0 —	0 — 0 —	0 — 10 +	15 + 0 —	(0) (—) (0) (—)	0 — 0 —	0 (—) 5 (+)
21	0 — 0 —	(0) (—) (0) (—)	0 — 0 —	0 — 0 —	0 — 0 —	0 — 0 —	0 — 0 —	(0) (—) (0) (—)	45 + 0 —	0 — 0 —	(0) (—) (0) (—)	92 + 0 —
22	0 — 0 —	(0) (—) (0) (—)	0 (—) 0 (—)	2 — 0 —	55 + 34 +	0 — 0 —	(0) (—) (0) (—)	0 — 0 —	0 — 0 —	0 — 0 —	13 + 0 —	0 — 0 —
23	(0) (—) (0) (—)	(0) (—) (0) (—)	(65) (+) (5) (—)	0 — 0 —	60 + 11 +	20 + 0 —	0 — 1 —	(0) (—) (0) (—)	30 + 0 —	0 (—) 0 (—)	6 (+) 0 (—)	0 — 0 —
24	(0) (—) (0) (—)	(0) (—) (0) (—)	25 — 0 —	(0) (—) (0) (—)	0 — 0 —	(0) (—) (16) (+)	1 — 10 +	0 (—) 0 (—)	3 + 0 —	0 — 0 —	13 + 2 —	(0) (—) (0) (—)
25	(0) (—) (0) (—)	(39) (+) (0) (—)	0 — 0 —	0 — 36 +	0 — 6 —	0 — 0 —	0 — 4 +	0 — 0 —	0 — 0 —	0 — 0 —	(0) (—) (0) (—)	(0) (—) (0) (—)
26	0 — 0 —	0 — 0 —	0 — 0 —	13 (+) 0 (—)	0 — 0 —	(0) (—) (0) (—)	0 — 2 +	(0) (—) (0) (—)	(0) (—) (0) (—)	(0) (—) (0) (—)	0 — 0 —	(0) (—) (0) (—)
27	(0) (—) (3) (+)	(0) (—) (0) (—)	0 — 0 —	14 + 0 —	0 — 0 —	12 — 0 —	0 — 0 —	0 — 0 —	0 — 0 —	0 — 0 —	0 (—) 0 (—)	0 — 0 —
28	(16) (+) (18) (+)	(0) (—) (0) (—)	(0) (—) (0) (—)	(0) (—) (2) (—)	(0) (—) (51) (+)	0 — 0 —	0 — 6 +	0 — 0 —	0 — 0 —	(0) (—) (0) (—)	0 — 0 —	(0) (—) (0) (—)
29	(60) (+) (1) (+)		9 (+) 0 (—)	(0) (—) (11) (+)	0 (—) 0 (—)	0 (—) 0 (—)	(0) (—) (0) (—)	0 — 0 —	0 — 0 —	0 — 0 —	5 + 0 —	(0) (—) (0) (—)
30	(16) (+) (0) (—)		0 — 0 —	(0) (—) (1) (—)	0 — 0 —	0 — 0 —	0 — 0 —	0 — 0 —	(3) (+) (0) (—)	0 — 0 —	0 — 0 —	(0) (—) (0) (—)
31	0 — 0 —		0 — 0 —		35 + 0 —		0 — 0 —	0 — 0 —		(0) (—) (0) (—)		42 + 0 —

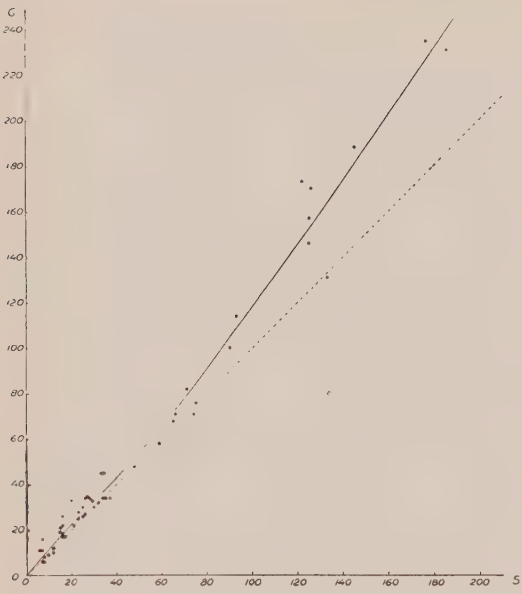


Fig. 1. Comparison between Greenwich Observatory sunspot areas G and Stockholm Observatory sunspot areas S.

servation was possible that day but the value has been obtained by interpolation. If this is the case for several days in succession the interpolation has been extended over several days. Zero means that no spot is situated in the day-sector after the central meridian and — that a long period of cloudy weather has made the determination impossible or very uncertain.

When the numbers in table I and table II are followed by a + sign, this means that the sunspot group is surrounded by bright *Hα* faculae according to spectroheliograms obtained the same day with our patrol spectroheliograph. A + sign within parenthesis means that no spectroheliogram has been obtained the same day but that the activity of the group is judged from spectroheliograms obtained before or after that date. A — sign after the numbers means that no noticeable bright *Hα* faculae are seen on our spectroheliograms and a — sign within parenthesis that the low activity of the sunspot group is judged from earlier or later observations. If neither a + sign nor a — sign is accompanying the number a long period of cloudy weather has made even an interpolation of the activity character impossible.

The spectroheliograph used for the patrol work is the same as that described in the previous work. It has been equipped with two rotating Anderson prisms so that the instrument can be used not only as a spectroheliograph but also as a spectrohelioscope.

As the sunspot areas determined by us are based on photographs obtained with a small instrument and in the violet light it has been interesting to compare our determinations with those made with a considerably larger instrument at the Greenwich Observatory. By the great courtesy of the Astronomer Royal we are able to present in fig. 1 a comparison between our sunspot areas (S) and those determined at the Greenwich Observatory (G). For the comparison a number of well defined groups appearing near the sun's central meridian in April and May 1948 have been selected.

According to the comparison presented in fig. 1 our sunspot areas are systematically somewhat smaller than the Greenwich values. The reason for this my be the smaller scale of our photographs and the different wavelength regions used. The scattering of the individual points is moderate, however, which indicates that our determinations may be used as representative values of sunspot areas. In fact some of the greatest deviations seem to be due to rapid changes in the size of the spots.

B. *Solar flares.* During 1950 and 1951 a number of flares of intensity 3 have been observed at the Stockholm Observatory with the patrol spectroheliograph. All these flares have been recorded after immediate announcements by the Royal Swedish Board of Telegraph of fade outs in the short wave communications. These observations are listed in table III together with some smaller flares observed by us accidentally.

C. *Filaments.* A special attention has been given to the appearence of large filaments as such objects sometimes produce ionospheric disturbances three or four days after the transit over the sun's central meridian. The large filaments are generally very outstanding objects on our *Hα* spectroheliograms. Those recorded by us in or near the first day-sector after the central meridian are listed in table IV and table V. The tables should be compared with tables I and II in order to judge which periods may be complete and which may not

be complete. For such days when no + or no — sign is given in table I and table II no information about the presence of filaments has been available.

D. *Polar maps for announcing the solar activity.* For the distribution twice weekly of reports on the general solar activity it has been found very appropriate to use polar maps of the sun of the type presented in fig. 2. The sun's north

Table III. Flares observed at the Stockholm Observatory

Day	First obs	Max	Last obs	Imp	Lat	Long
1950	h m	h m	h m			
May 3	9 56	9 56	10 28	3	14N	1E
5	10 07	10 10	10 44	2	12N	29W
6	13 38	13 39	13 41	2+	12N	47W
10	10 50	11 14	11 22	2—3	6N	26E
Dec. 26	10 31	10 31	10 38	1	12N	18E
1951						
May 18	10 45	10 47	11 19	3	15N	30W
June 8	15 46			1—2	14N	51E
13	06 42			2—3	10N	15W
13	07 48			1	13S	75E
Sept 3	13 08	13 15	14 10	3	5N	30E

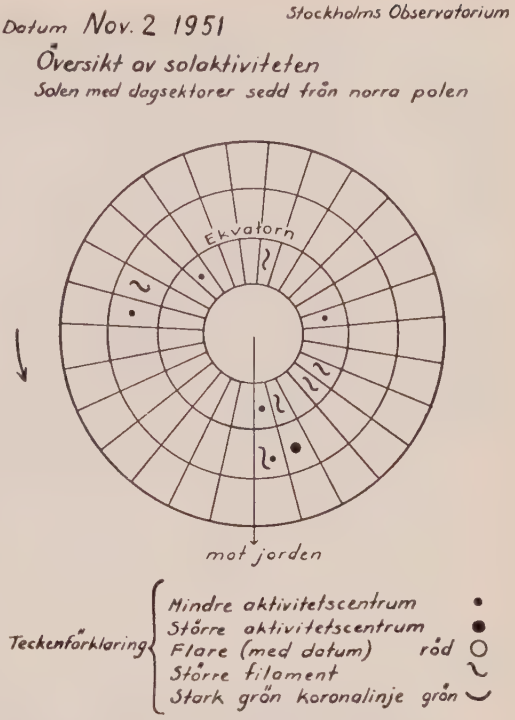


Fig. 2. Stockholm Observatory solar map for Nov. 2, 1951 giving the distribution of activity centres (spots) and filaments. The arrow is pointing towards the earth.

Table IV (1950)

Jan.	Febr.	March	April	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.
5	(16)	1	4	5	6	1	4	2	2	(2)	(1)
29	25	10	7	8	8	2	9	5	(4)	(9)	9
	28	15	(11)	10	(11)	5	11	7	5	12	(15)
		16	20	(15)	15	7	12	8	(14)	14	(29)
		23	21	18	16	8	14	10	18	(17)	
		30	22	25	17	13	20	(16)	23		
			(24)		18	15	24		27		
			30		23	18	29				
						19					
						(23)					
						25					
						26					
						29					
						31					

Table V (1951)

Jan.	Febr.	March	April	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.
9	19	3	17	5	12	6	6	2	2	(1)	6
	(28)	(4)	(30)	14	(14)	(9)	9	4	(13)	15	(14)
		(8)			28	12	(12)	5	17	26	22
		(10)				18	19	(9)	22	29	(26)
						21	(26)	15	27		
							30	28	29		
									30		

pole is used here as origin and 27 equal sectors are given, each one corresponding to the synodic rotation of the sun in 24 hours. The first zone is used for characterizing the solar phenomena north of the equator and next zone is used for giving the solar phenomena south of the equator. Different phenomena such as spots, flares, filaments and a strong intensity of the green corona line are indicated by means of different symbols. In this way a very clear picture is obtained of the whole sun and it is easy to see when old phenomena may be expected to appear again at the east limb.

During such periods when the Stockholm Observatory coronagraph has been in regular use the regions of high corona intensity have been obtained from our own recordings. For the information of different geophysical institutions in Sweden we have also indicated high corona intensity values obtained from the Ursigrams. The postcard contains predictions for ionospheric disturbances as well.

For our own work a big map of the same kind is used. This map can be turned around its axis so that the proper setting is obtained for each day. The device is in a way more simple than the globe used in some astronomical institutions for the same purpose.

Stockholm Observatory, Saltsjöbaden June, 1952.

Localization of Aurorae with 10 m High Power Radar Technique, using a Rotating Antenna

By GÖSTA HELLGREN and JOHAN MEOS
Research Laboratory of Electronics, Chalmers University of Technology, Gothenburg

(Manuscript received 5 August 1952)

Abstract

The paper describes the 10 m high power recorder with a rotating antenna that is used since May 1951 for the localization of aurorae at the Radio Wave Propagation Laboratory of the Kiruna Geophysical Observatory (67.8° N, 20.5° E). Continuous observations during the time May 1951—March 1952 have disclosed periods of auroral activity. The preliminary results from these observations indicate that there is a good correlation between the auroral activity, the magnetic activity, and the appearance of the N1-layer, a special type of sporadic E ionization often appearing in connection with magnetic bays and supposed to be caused by the same ionizing agent as the aurora. The distribution in range and bearing of the recorded aurorae agrees with the simple theory that most of the radio wave scattering comes from those points where the radar beam is perpendicular to the surface of the auroral discharges. The calculated height distributions of the reflection centres have maxima around 120 km.

1. Introduction

In order to extend the study of the general features of the polar ionosphere and other associated phenomena it was early planned by the Chalmers Research Laboratory of Electronics to study long distance scattering of radio waves from the Aurora Borealis (Rydbeck, 1949). Since May 1951 the Radio Wave Propagation Laboratory of the Kiruna Geophysical Observatory (see fig. 1) has therefore been equipped with an aurora recorder of radar type operating on 10 m wavelength (frequency 30.3 Mc/s) using a rotating antenna. A great number of very interesting echoes have been received. Their distribution in range and bearing and the correlation with visually observed aurorae strongly support the theory that echoes are caused not by an ordinary type of ionosphere (ordinary or sporadic E) but by reflections from auroral discharges.

Scattering of radio waves from the Aurora Borealis has earlier been reported by Harang (1940) on a frequency of 40 Mc/s and by Lovell, Clegg, and Ellyett (1947) on 46 Mc/s. During 1949 Aspinall and Hawkins (1950) identified a number of aurora formations in the range 400—900 km by the characteristic echoes which were produced on 72 Mc/s.

This paper describes our 10 m aurora recorder and presents the observations and the preliminary results obtained during the period May 1951—March 1952.

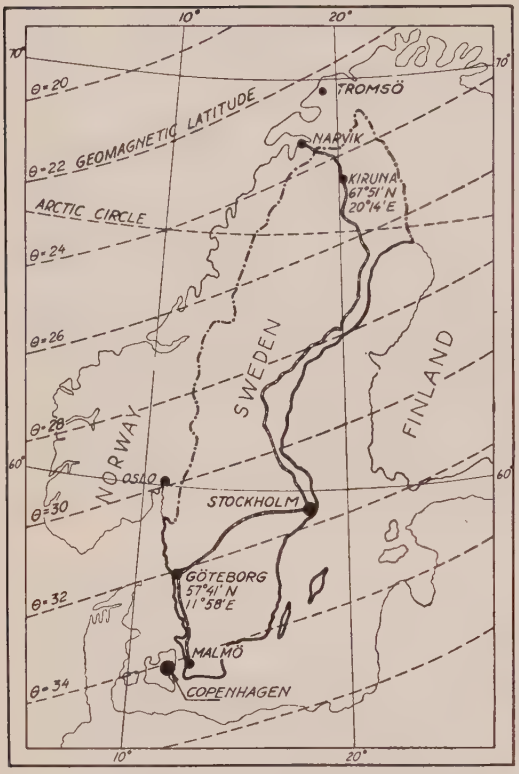


Fig. 1. Location of Kiruna Geophysical Observatory.

2. Recording equipment

The equipment is principally the same as that used at our Råö Observatory in the vicinity of Gothenburg for the recording of meteor trails. Power is transmitted and received by the same antenna. The received echoes are shown on an intensity modulated cathode ray tube display of PPI-type, which is photographed. The transmitter is a quite conventional self-pulsed tuned-grid-tuned-plate oscillator with 100 kW peak pulse power, 40 μs pulse width, and a pulse repetition rate of 50 c/s. The rotating antenna, a three-element Yagi, is mounted on the roof of one of the observatory buildings (see fig. 2). It is elevated 30° above the horizontal and has an estimated gain of 9 db compared with that of an isotropic radiator. The receiving unit of the recorder (fig. 3) consists of a modified Hallicrafter receiver, type R44/ARR-5, and a 10 cm radar SO-13 indicator. The smallest signal detectable on the cathode ray tube is equivalent to 3×10^{-14} W at the receiver input. The antenna and the synchronized beam on the scope make two rotations per minute. The camera is open for exposure for 30 seconds, i. e. the time of one complete revolution of the antenna. It can be run automatically either continuously, taking one picture of each rotation, or intermittently, taking only 12 pictures every full hour.

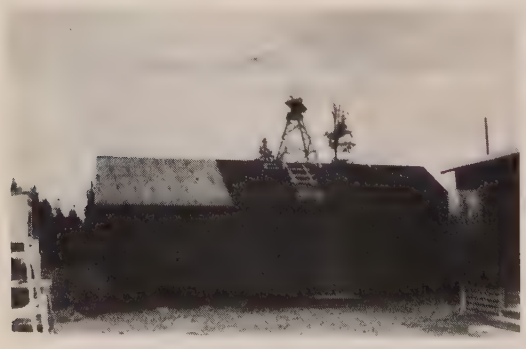


Fig. 2. Rotating antenna of 10 m aurora recorder.

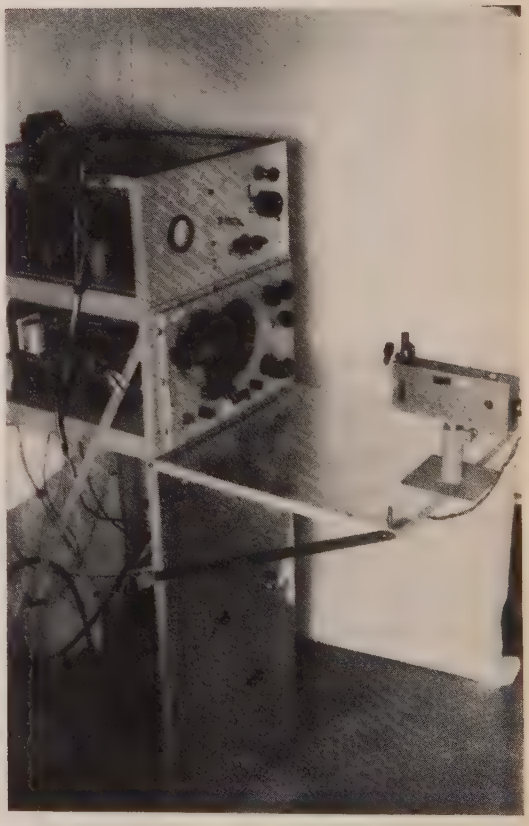


Fig. 3. Receiving unit of 10 m aurora recorder.

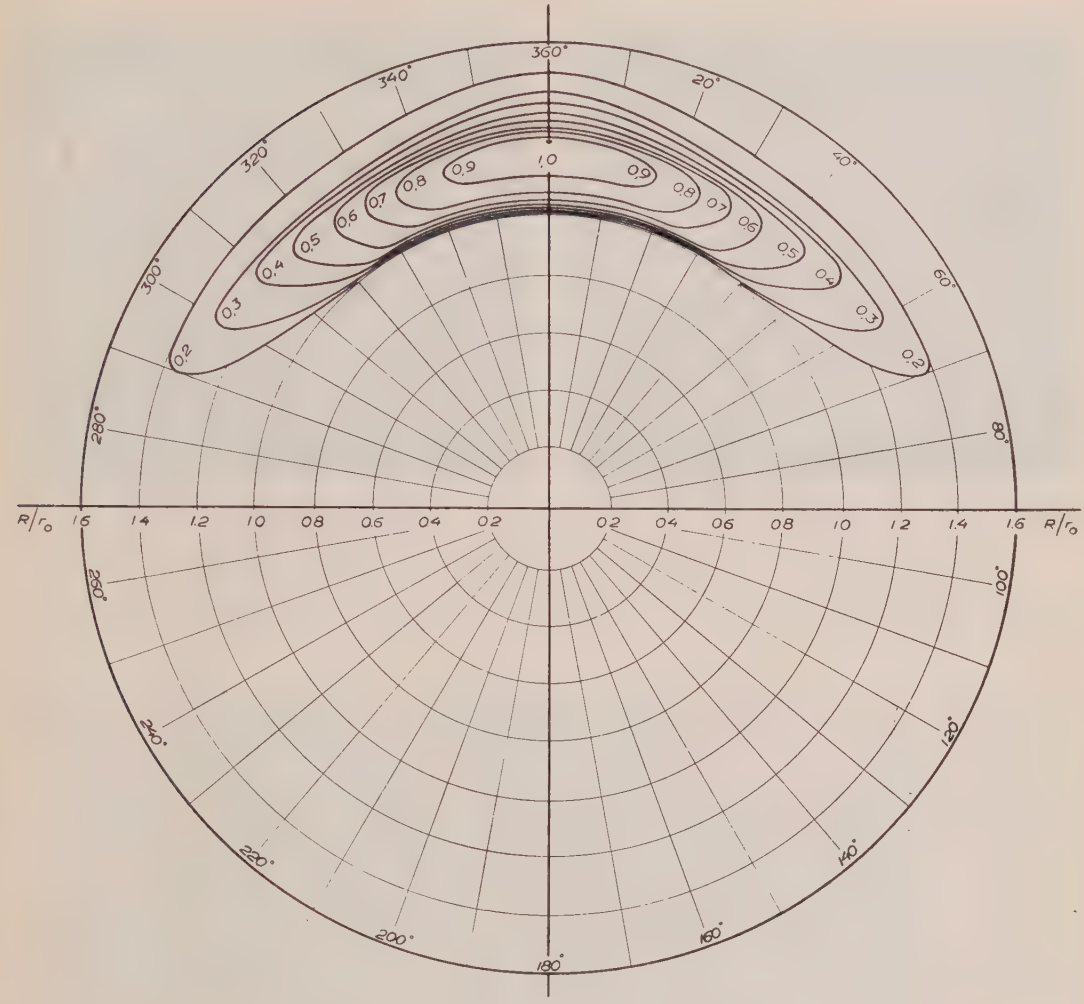


Fig. 4. Theoretical radio echo from an auroral curtain.

3. Types of echoes

Roughly speaking, the great variety of auroral echoes can be classified into one or the other of two different types, diffuse or discrete. The diffuse echo must be due to reflections from an extended irregular surface or a large heterogeneous region with a random distribution of electron density. The discrete echo, on the other hand, must originate from a regular surface containing a relatively homogeneous electron distribution. Therefore, as already done by ASPINALL and HAWKINS (1950) it is reasonable to associate a diffuse echo with an auroral glow or a curtain, and a discrete one with a streamer. In the following we will use the terminology, curtain, for a diffuse echo, and streamer, for a discrete one. Most of the reflected energy must come from those points of the curtains and streamers where the radar beam is perpendicular to the corresponding surfaces. For a certain locality the most favourable direction of reflection is the direction of the magnetic north, if one assumes the auroral discharges to take place on surfaces generated by the magnetic field lines which pass through curves on the earth's surface of constant horizontal force.

Applying the theory of radio wave scattering from a turbulent medium (BOOKER, GORDON, 1950) to the case of reflection from an auroral curtain we get a diffuse echo on the indicator screen like that illustrated on fig. 4, when the

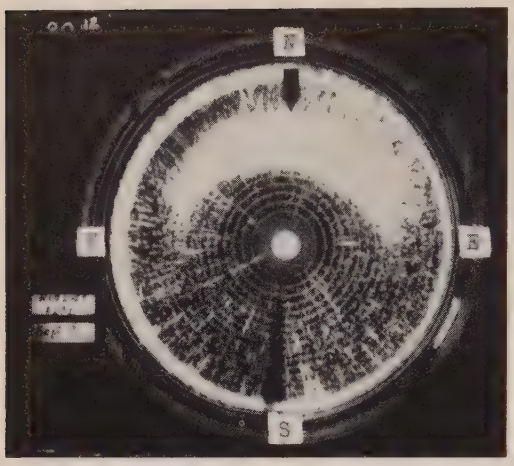


Fig. 5. Typical radio echo from an auroral curtain.

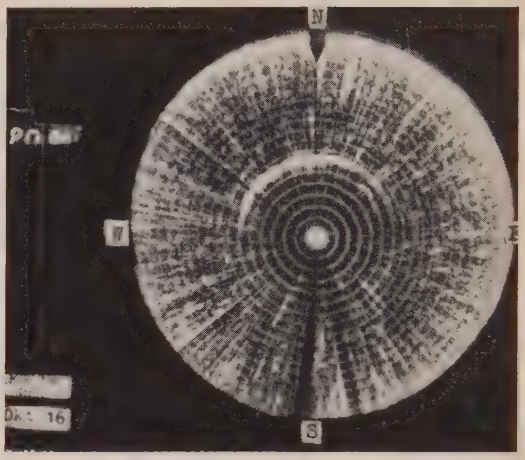


Fig. 6. Typical radio echo from an auroral streamer.

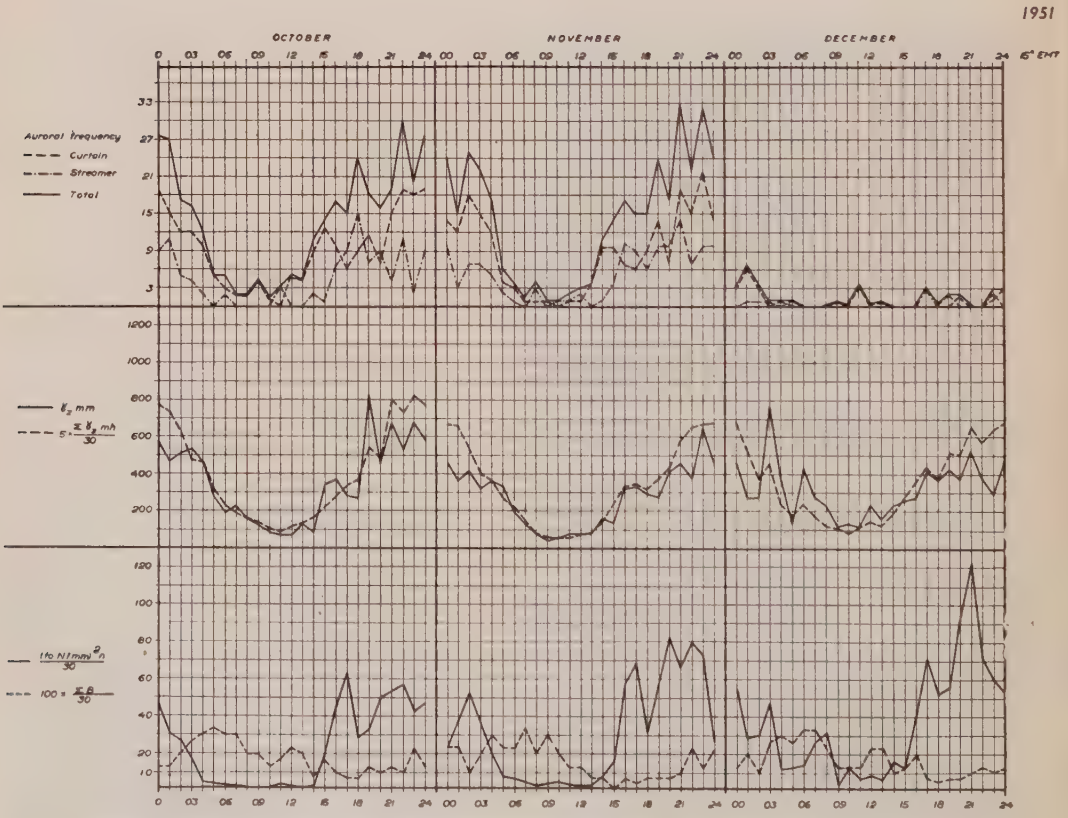


Fig. 7. Hourly variation of the auroral and the magnetic activities, the intensity of the N1-reflections, and the appearance of polar blackouts during October—December 1951.

antenna rotates. We have here assumed the curtain to be an infinitely extended plane sheet with a thickness of 0.25 R_0, where R_0 is the minimum range to the sheet, and the antenna to have constant gain throughout an angle of 90°. Further, we have assumed that the scale of turbulence is large compared with the wavelength, and that the mean square deviation of the dielectric constant of the layer is small. If, instead, the reflections had come from an auroral streamer, supposed to be a circular cylinder (with a diameter of the order of 2 km) containing a uniform electron density the echo picture would have been a uniformly bright, very narrow 90° arc of a circle. Figs. 5 and 6 show the two different echo types, as obtained by our Kiruna recorder on September 22nd, 2012 15°EMT and on October 16th, 2001 15°EMT respectively. The arrow on the top indicates the north, east is to the right and west to the left. The difference in range between two concentric circles is

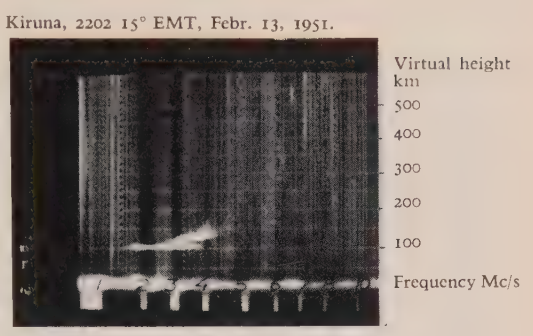


Fig. 8. Record showing simultaneous reflections from sporadic E-layer (100 km height) and N1-layer (115 km height).

50 km and the maximum range 1,000 km (this could be increased if required up to 2,000 km but is ordinarily quite sufficient).

4. Preliminary recorded results

During the time May 1951—March 1952 the aurora recorder ran day and night with the

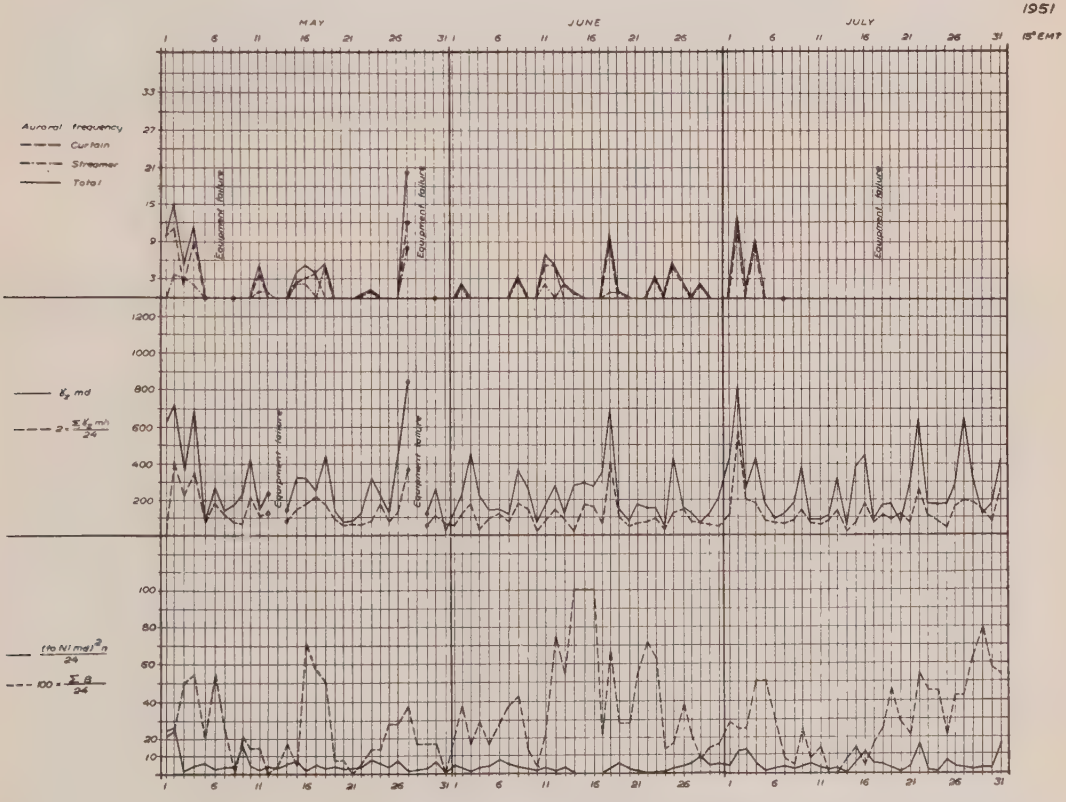


Fig. 9. Day to day variation of the auroral and the magnetic activities, the intensity of the N1-reflections, and the appearance of polar blackouts during May, June, and July 1951.

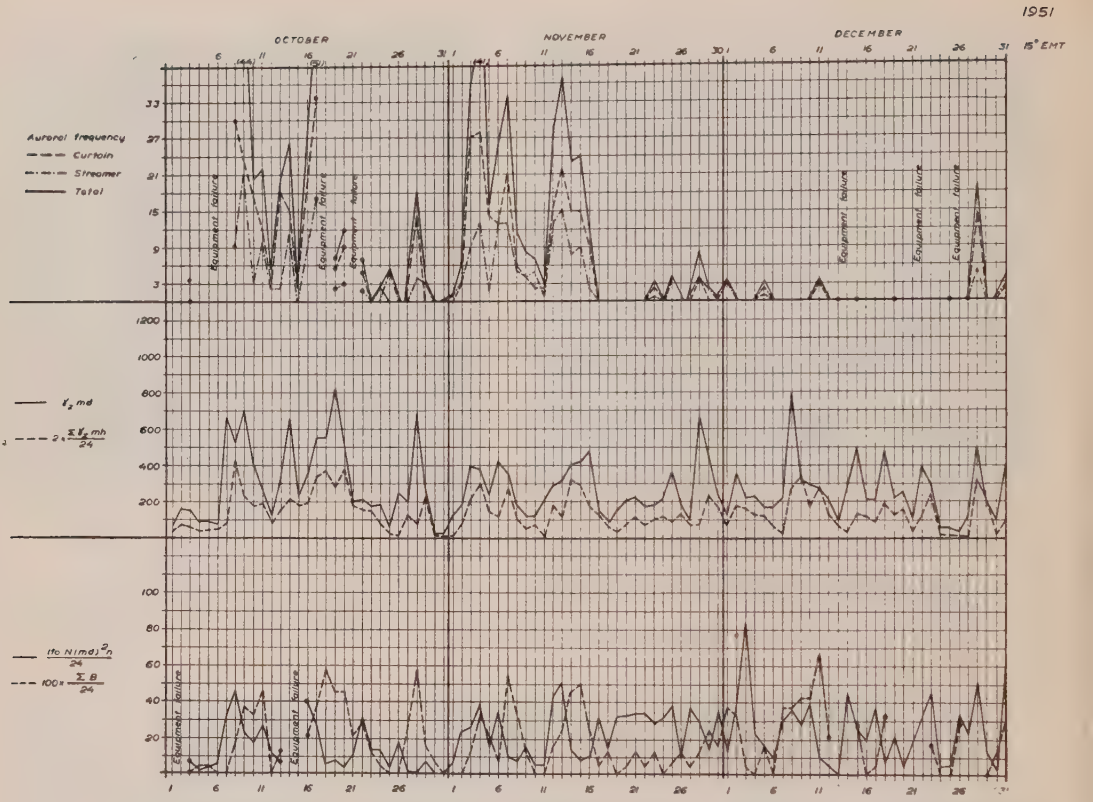


Fig. 10. Day to day variation of the auroral and the magnetic activities, the intensity of the N1-reflections, and the appearance of polar blackouts during October, November, and December 1951.

exception of a period of relatively small auroral activity in August and September 1951 when part of the installation had to be made more permanent. Normally, in order to save film, only 12 consecutive pictures were taken during every hour, but in the case of special auroral activity, and at least once a week, each sweep was photographed for a period of 24 hours, when, also, the ordinary panoramic ionospheric recorder was in continuous operation. The hourly recordings are quite sufficient to permit statistical investigations of the aurora, but the continuous films, of course, give a better and a dynamic picture of the whole phenomenon, and are of special value for a detailed comparison with the ionospheric soundings and the magnetic records.

To interpret the hourly records one characteristic echo picture out of the 12 obtained has been analyzed with regard to the appearance of diffuse echoes (curtains) and discrete echoes (streamers). According to its strength each echo has been given a weighted number in the following way:

1 = aurora of small intensity
2 = aurora of moderate intensity
3 = aurora of strong intensity

(The maximum auroral activity, or frequency, of either kind is consequently during 24 hours $24 \times 3 = 72$). Of course, many objections could be raised to such a procedure; the personal errors are important in the reading of the records, no consideration has been given to the influence of distance on the echo intensity and so on. The method is justified only by the fact that it forms a basis for a preliminary statistical investigation of the observations.

The variation throughout the day of the auroral activity, as shown by the hourly records, indicates a maximum activity around midnight and a minimum around noon. Fig. 7

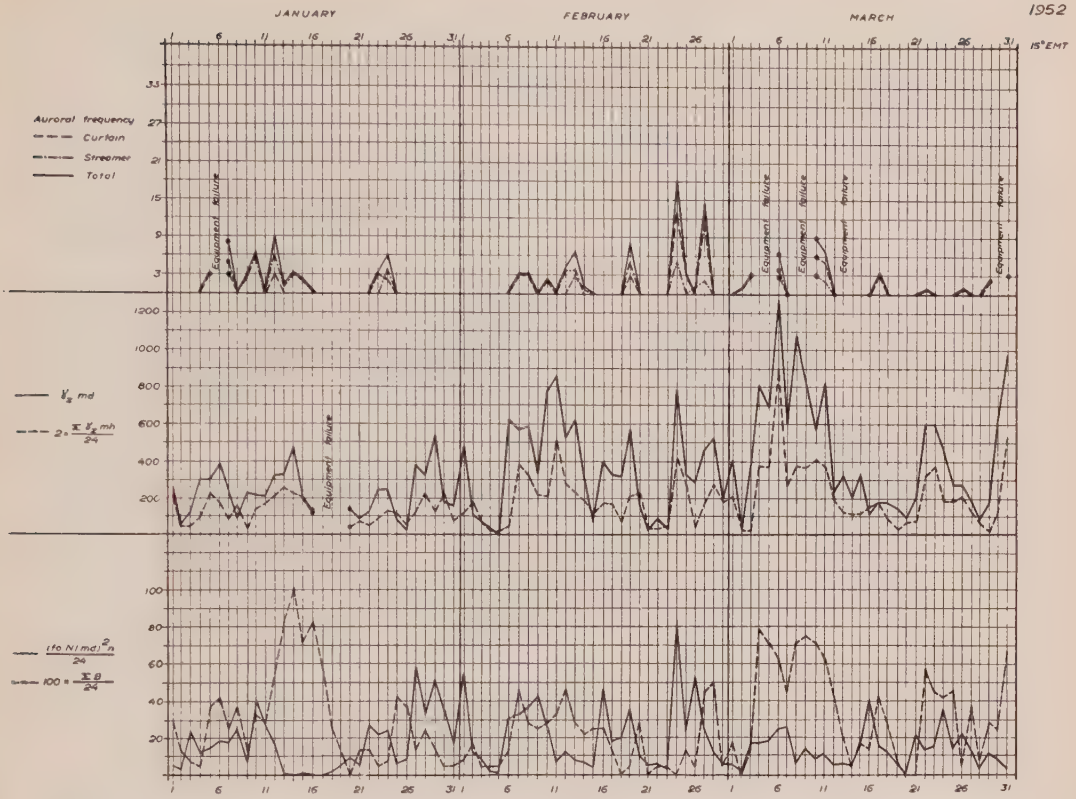


Fig. 11. Day to day variation of the auroral and the magnetic activities, the intensity of the N1-reflections, and the appearance of polar blackouts during January, February, and March 1952.

shows the curves obtained for October and November 1951, the months of highest auroral activity, and for December 1951, a month with relatively low activity. The curves are typical. On the same graphs the hourly variation of the magnetic activity as well as the intensity of the so called N1-reflections and the number of polar blackouts has been plotted, as given by our Kiruna magnetograms and ionospheric soundings. $\gamma_z mh$ is here defined as the maximum departure in γ of the magnetic z-component from undisturbed value during a certain hour, and $\gamma_z mm$ is the maximum value of $\gamma_z mh$ for the whole month. $foN1mm$ is the highest frequency for a certain hour during the whole month on which echoes from the N1-layer have been observed, and n is the number of times that N1-echoes have been obtained during the month ($n \leq 30$). ΣB is the total number of

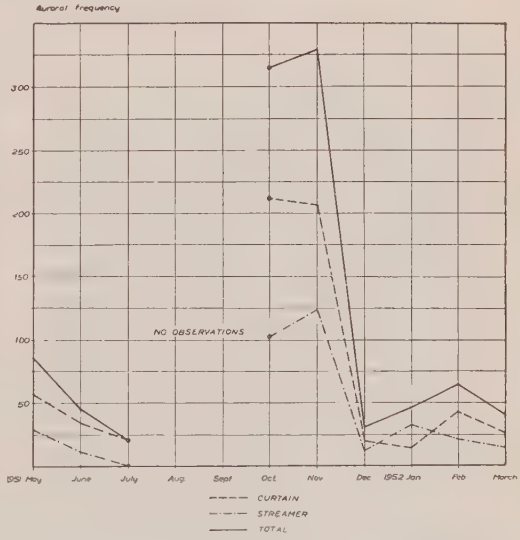


Fig. 12. The monthly variation of the auroral activities during the time May 1951—March 1952.

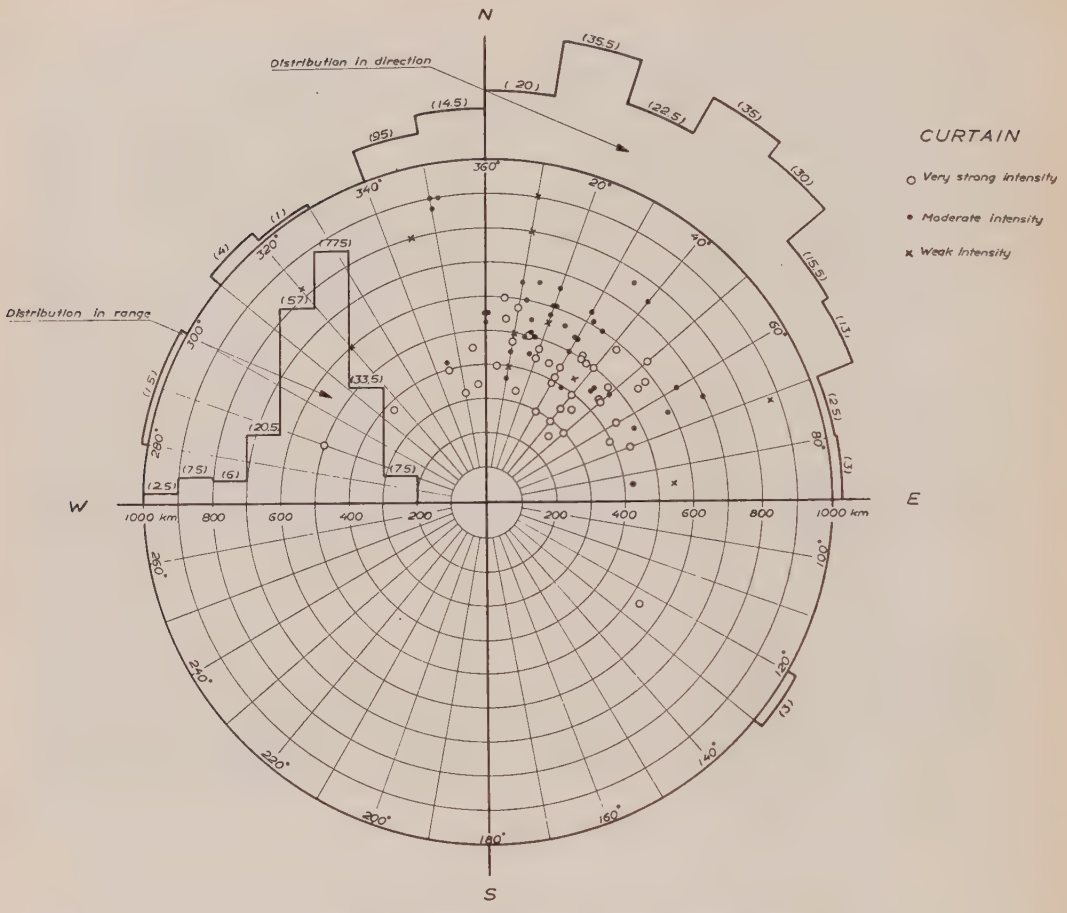


Fig. 13. The reflection centres of auroral curtains during October 1951.

blackouts observed during a certain hour for the whole month ($\Sigma B \leq 30$).

The N1-layer is a special type of sporadic E-ionization often appearing in polar regions and always in connection with magnetic bays and supposed to be caused by the same ionizing agent as the aurora. The hf-trace of N1 looks quite similar to that of the ordinary E-layer except for the fact that the N1-layer seems to be thicker and often shows complicated splitting near the penetration frequency (see fig. 8). For more detailed information about the N1-reflections the reader may be referred to LINDQUIST (1950).

The polar blackout is a certain type of "no echo" condition, which occurs in polar regions and shows no apparent connection with solar flares. The disappearance of reflections from all ionospheric layers during a blackout is believed to be due to an abnormally high absorption in a region below the 100 km level. This increased absorption may be caused by the impact of some ionizing agent in the ionosphere. The N1-reflection appears before and after almost every blackout (LINDQUIST, 1951).

As can be seen from the graphs there exists a good correlation between the auroral frequency, the magnetic activity, and the intensity of the N1-reflections. It is very interesting to note that the curve of the blackouts has the same form as the curves of the other phenomena but is delayed about 6 hours.

The day to day variation of the recorded auroral activity is given on figs 9, 10, and 11 for the whole period of observation, May

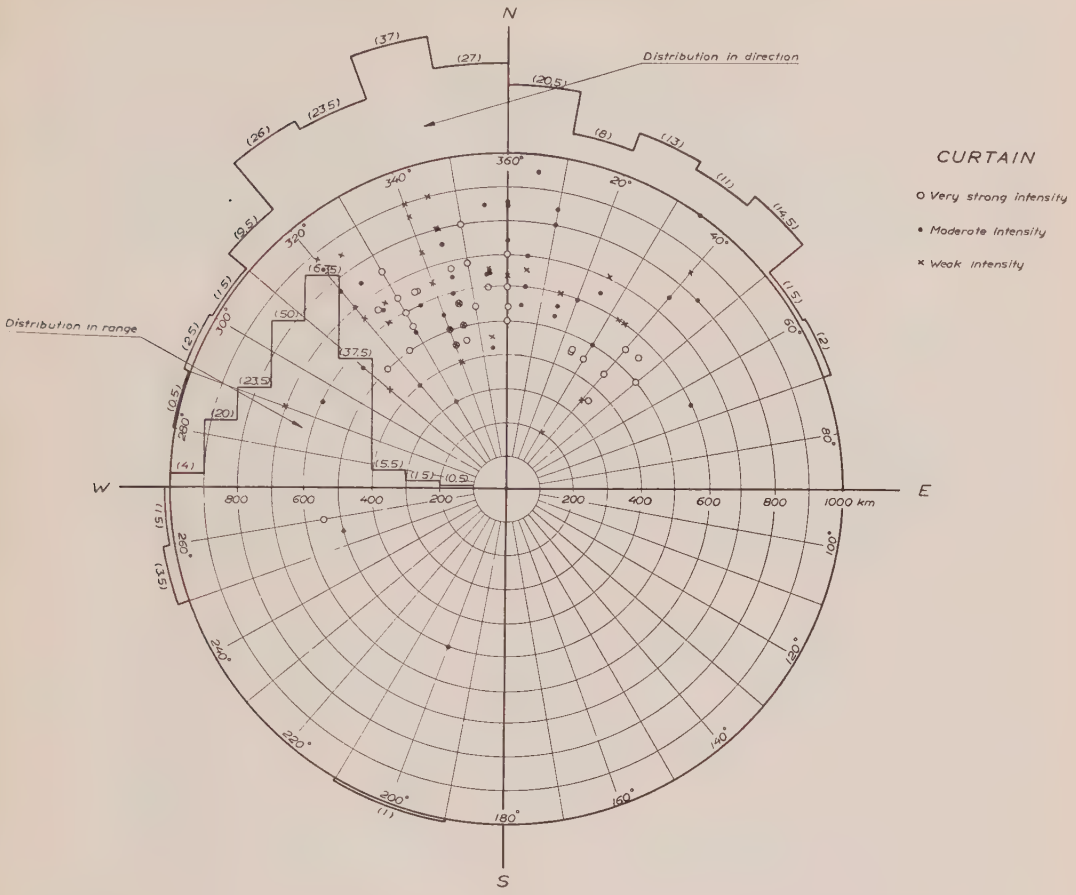


Fig. 14. The reflection centres of auroral curtains during November 1951.

1951—March 1952. On the same graphs we find the corresponding curves for the magnetic activity, the intensity of N1-reflections, and the appearance of polar blackouts. $\gamma_z md$ is here the maximum departure from undisturbed value of the magnetic z-component during the day, and $foN1md$ is the highest frequency during the day on which echoes from the N1-layer are observed. n is the number of times during the day that $foN1$ has been obtained ($n \leq 24$). ΣB is the total number of hours during the day on which polar blackouts occurred ($\Sigma B \leq 24$). There is a good correlation between all the curves and especially between those of the auroral activity and the appearance of polar blackouts. No significant difference exists between the two kinds of radar echoes, curtains and streamers.

The monthly distribution of the recorded auroral frequencies has a very marked maximum in October and November 1951 and a second but smaller maximum in Februari 1952 (see fig. 12). This is in good agreement with what is known about aurorae from visual observations during many years.

During the time of observation the highest auroral frequency always appeared in the bearings northeast to northwest and very seldom in the south. This well agrees with the simple theory that the most probable direction of reflection is the direction of the magnetic north. To show typical cases we have plotted in polar coordinates on figs 13, 14, 15, and 16 all the centres of reflections obtained during October and November 1951 for auroral curtains and streamers respectively.

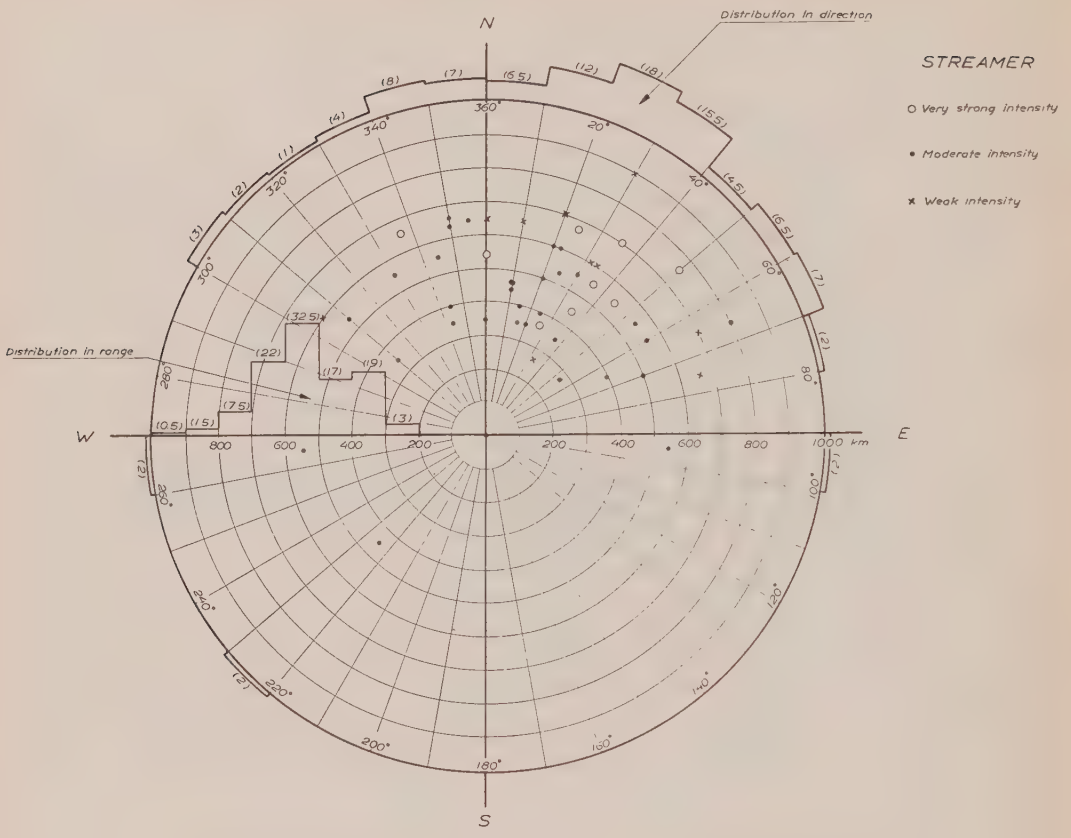


Fig. 15. The reflection centres of auroral streamers during October 1951.

It is interesting to note that the optimal direction of reflection was constantly north to northeast during all the months except November when it changed to north-northwest. As for the distribution in range, the reflection centres of curtains have a maximum in the interval 500—600 km, and those of streamers, in the interval 550—650 km, as can also be seen on the same figures.

From the direction and range measurements it is possible to calculate the height distribution of the reflection centres if one assumes the maximum reflected energy to come from those points where the radar beam is perpendicular to the reflecting surface. These surfaces are supposed to be circular cylindrical surfaces parallel with the lines of magnetic force (streamers), or surfaces generated by the lines of force passing through the curves on the earth's surface of constant magnetic horizontal intensity (curtains). Such calculations have been made for all the reflection centres obtained within an angle of $\pm 45°$ from the direction of the magnetic north (at Kiruna: magnetic declination 2°E, inclination 77°). The results for October and November 1951 are given in fig. 17. The height distributions have maxima of about 120 km and thus they fall in the region of maximum luminosity of visible aurorae. (CHAPMAN, BARTELS, 1940).

The continuous radar recordings of the aurora give an indication of the comparatively rapid appearance and disappearance of the discharges, as well as the swift movements of the aurora, or rather the swift displacement of the ionizing agent causing the aurora. This rapid variation is illustrated by the collection of pictures reproduced on fig. 18. The pictures show an aurora of a complex type with both curtains and streamers, lasting from 2228 to 2245 15°EMT, October 26th 1951. A

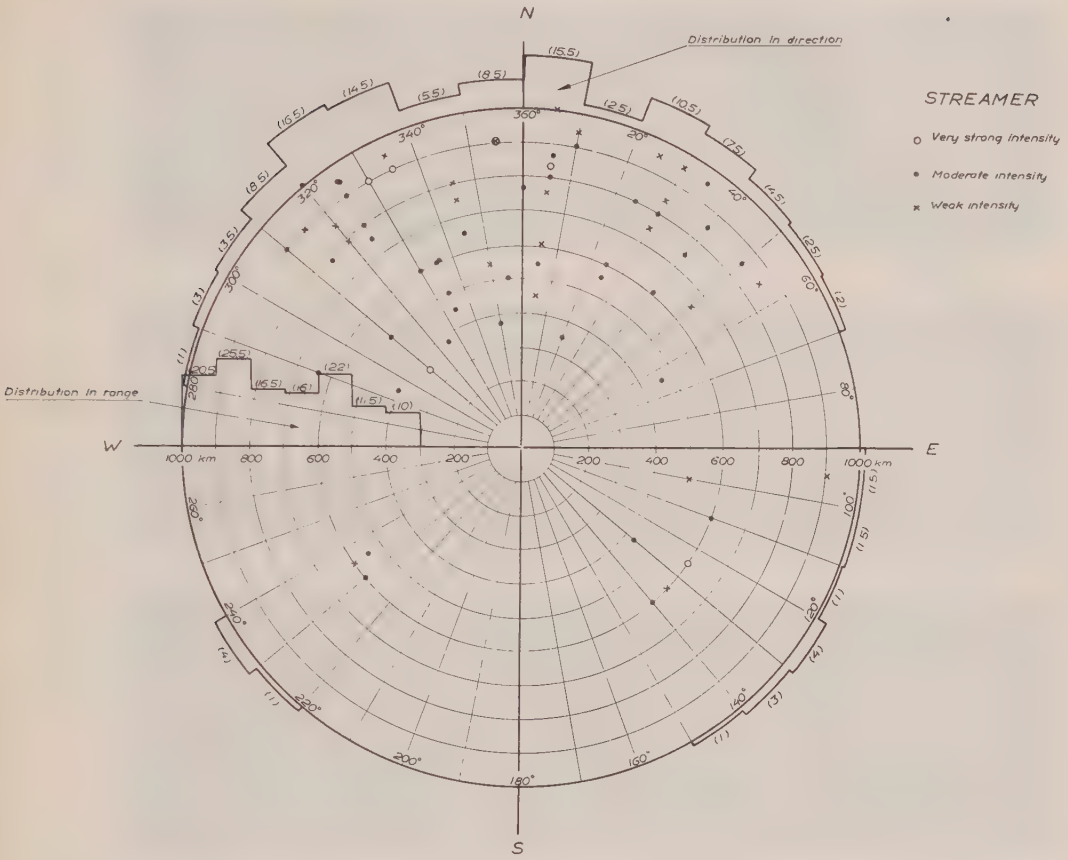


Fig. 16. The reflection centres of auroral streamers during November 1951.

study of the continuous recordings gives times of duration of about 240 and 100 seconds for diffuse and discrete echoes respectively. The longest time observed during a representative period in October 1951 was 420 seconds for curtains and 240 seconds for streamers. Echo times shorter than 30 seconds occurred very seldom.

From the amount of the reflected energy Aspinall and Hawkins (1950) have calculated the typical electron density of a streamer to be approximately 6×10^5 electrons per cc, supposing the streamer to be a circular cylinder (radius 1,000 m) enclosing a region of uniform electron density. An estimate on the same basis that we have made in some cases where the amount of the reflected energy could be determined with enough accuracy, gives densities of the order of 2×10^6 electrons per cc as typical for streamers. This density value is

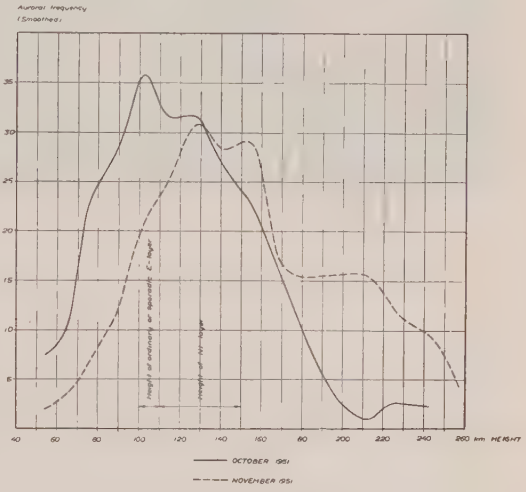


Fig. 17. Height distribution of reflection centres within an angle of $\pm 45°$ from the direction of the magnetic north of Kiruna during October and November 1951.

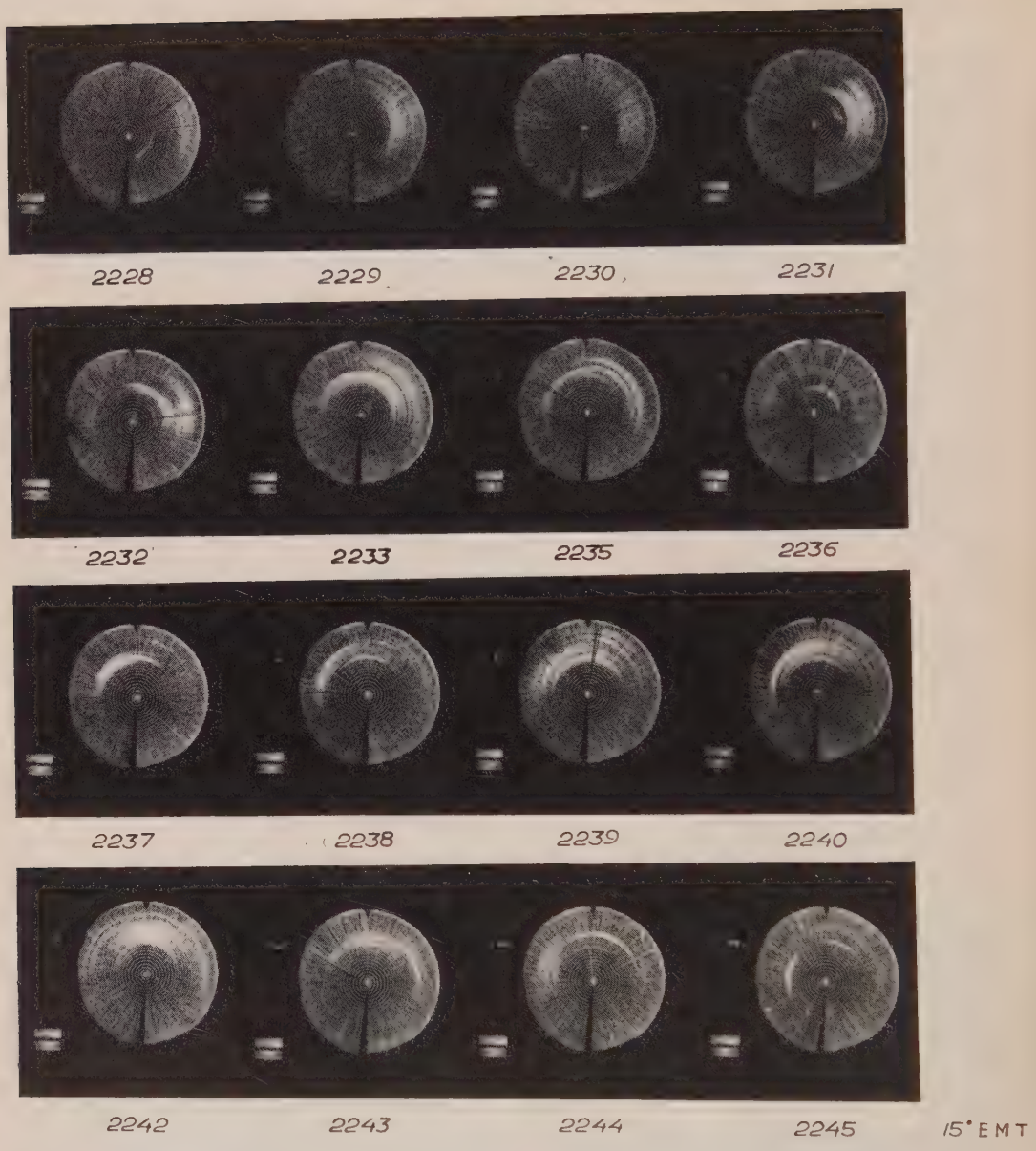


Fig. 18. A series of records from an aurora appearing between 2228 and 2245 15° EMT, October 26, 1951.

sufficient to explain the rapid fluctuation of the echo pattern mentioned above. Solving the inhomogeneous Riccati equation, describing the simplest type of production and recombination of electrons

$$\frac{dN}{dt} = q\,(t) - \alpha N^2,$$

where

N = electrons per cc
q = production of ions per cc and second
α = effective recombination coefficient
t = time,

for the case of a sudden disappearance of the ionization agent $(q\,(t) = q_0,\ t \leq t_0;\ q\,(t) = 0,\ t > t_0)$ gives as a solution

$$N = \frac{N_0}{1 + \alpha N_0 (t - t_0)}; \quad t \geq t_0.$$

Taking N_0, the initial value of the density, equal to 2×10^6 and choosing the value of the recombination coefficient found by RYDBECK (1946) for the ionosphere at the height of 100 km, $\alpha = 1.1 \times 10^{-8}$, we get an idea of the relaxation time τ by putting $\alpha N_0 \tau = 1$. This gives us $\tau \approx 45$ seconds, which is in good agreement with the observations.

5. Concluding remarks

Concurrently with the radar observations of the aurora visual observations were made. Because of unsuitable weather conditions the material is incomplete and a comparison between the results of the two observations is therefore very difficult. However, it turns out that during the months of highest auroral activity radar echoes were practically always seen on the scope when visual observations of the aurora were made, but often a simultaneous radar and visual aurora did not agree completely with respect to direction and character. There is little doubt, however, that the observed radio scatterings are due to reflections from auroral discharges. The distribution in direction and range of the echoes is the strongest evidence for the fact that the echoes could not come from the ionosphere or be any form of backscatter from ground via the ordinary or sporadic E-layer, the N1-layer, or the F_2-layer.

Acknowledgements

The work reported here has been possible through the generous grants given by the "Knut och Alice Wallenbergs stiftelse" and the "Wilhelm och Martina Lundgrens vetenskapsfond", to which the authors are greatly indebted. The authors thanks are also due, and are cordially extended to the Director of the Research Laboratory of Electronics, Professor O. E. H. Rydbeck, for his advice and support on many occasions. Finally, we wish to thank Dr. R. Lindquist, who supervised the development of the recording equipment and Messrs. O. Perers and B. Johanson, who helped during the observations.

REFERENCES

ASPINALL, A., and HAWKINS, G. S., 1950: Radio echo reflections from the Aurora Borealis. *J. Brit. Ast. Ass.*, **60.**

BOOKER, H. G., and GORDON, W. E., 1950: A theory of radio scattering in the troposphere. *Proc. I. R. E.*, **38**, No. 4.

CHAPMAN, S., and BARTELS, J., 1940: *Geomagnetism*, I. Oxford.

HARANG and STOFFREGEN, 1940: *Sonderdruck aus Hochfrequenztechnik und Elektroakustik*, **55.**

HERLOFSON, N., 1947: Interpretation of radio echoes from polar auroras. *Nature* **160.**

LINDQUIST, R., 1950: A survey of recent ionospheric measurements at the ionospheric and radio wave propagation observatory at Kiruna. *Arkiv för geofysik*, **1**, Nr. 11.

— 1951: Polar blackouts recorded at the Kiruna Observatory, *Trans. Chalmers Univ. of Technology*, **103.**

LOVELL, A. C. B., CLEGG, J. A., and ELLYETT, C. D., 1947: Radio echoes from the Aurora Borealis. *Nature* **160.**

RYDBECK, O. E. H., 1946: Chalmers solar eclipse expedition 1945. *Trans. Chalmers Univ. of Technology*, **53.**

— 1949: The ionospheric and radio wave propagation observatory at Kiruna, 67° 50′ N, 20° 5′ E. *Report No. 11, Res. Lab. of Electronics, Chalmers Univ. of Technology.*

Forty-Year Sea Level Pressure and Sunspots

By GLENN W. BRIER
U. S. Weather Bureau

(Manuscript received 15 June 1952)

Abstract

The 40-½ years of mean montly sea level pressure maps based on the Historical Map Series have been typed according to numerical indices that indicate to what extent each particular monthly map resembles a typical or normal winter or summer map on one hand, or a typical spring or fall map on the other hand. These indices are easily determined from the average pressure of certain key areas and are quite effective in separating the winter from the summer months and the spring months from the fall months. The persistence in these indices and their relation to the sunspot cycle has been investigated.

1. Introduction

In 1947 the writer (Brier 1947) in reporting upon an investigation of the pressure data from the Historical Daily Map Series (1943—44) found indications that there had been a considerable decrease in the total mass of air over the Norther Hemisphere between 1899 and 1939. The trend toward lower surface pressure was most pronounced in the higher latitudes and a further investigation (Brier 1948) disclosed that the trend existed in all months of the year. Ahlmann (1948), Kincer (1946) and others have reported that during this same period there seemed to be a general warming of the Northern Hemisphere, particularly in the higher latitudes and in the winter months. Because of general interest in the question of recent climatic fluctuations and the suggestion of Willett (1949) that irregular solar activity is the most probable cause of climatic fluctuations, it was decided to examine the variations in the pressure data in relation to known solar effects, such as the seasonal change in the position of the sun, and to other possible effects that might be associated with an index of solar activity such as sunspots.

2. The Seasonal Changes

Regarding the seasonal changes in the distribution of pressure, there are two well-known and established facts. First, in the Northern Hemisphere winter there is an excess of air relative to the summer. This is shown by the annual cycle of monthly values of the average sea level pressure north of Latitude 20° N when plotted for each month from January, 1899 to June, 1939, in Figure 1. These data are based on the series of Mean Monthly Pressure Maps (1946 b) for the period. Since they are based on reduced sea-level pressures rather than on actual station pressures they do not represent the best possible estimate of the total mass of air. However, some computations made by the author indicate that the

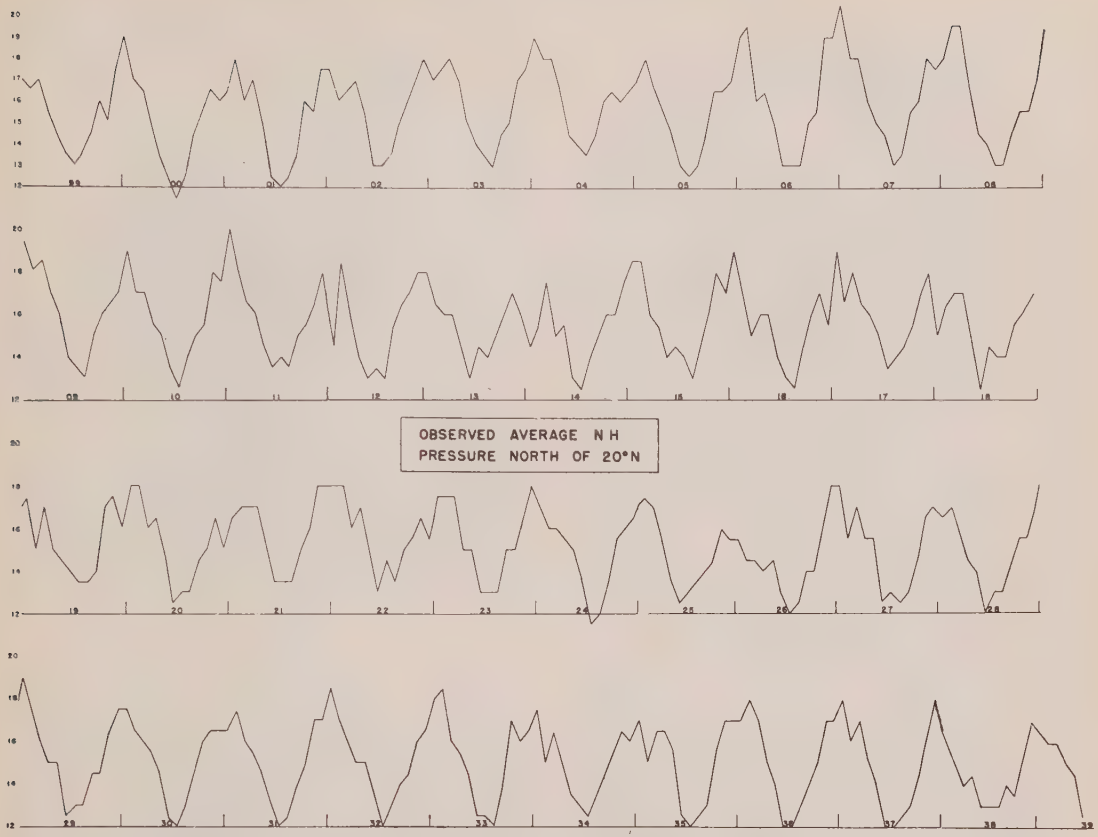


Fig. 1. Monthly values of the average sea-level pressure (in mb minus 1,000) north of latitude 20° N for the period 1899 to 1939. Numbers along the abscissa are years (without the first two numbers) plotted in the middle of the year.

effect of the artificial reduction to sea level is not a serious one for the purpose of this study and the qualitative picture given by Figure 1 is essentially correct.

The second well-known fact is that the relative distribution of air over the land and water surfaces of the hemisphere is, in general, opposite in winter and summer. This difference, a seasonal effect, is illustrated in Figure 2, which shows the normal difference in pressure between January and July. This map was obtained from the Normal Weather Maps (1946 a) which also were based on the 40 ½ years of sea-level pressure data. An additional chart, Figure 3, shows the difference in pressure between October and April and indicates that for most of the hemisphere the seasonal march of pressure is not symmetric to January and July.

3. Computation of Indices

Since Figure 2 indicates that the average difference in pressure between January and July is of greater magnitude, and of opposite sign, in some regions than in others, it would seem reasonable to suppose that a few critical or "key" areas could be chosen from which numerical pressure indices could be computed that would show a good discrimination between individual months of January and July. The author has shown (Brier, 1948) how monthly mean pressure maps might be typed or classified according to a few simple pressure indices chosen on the basis of areas showing maximum monthly change in surface pressure. An index, say W, might thus be computed for each of 486 months in the series, indicating whether a particular map was more like a winter map or a summer map. A

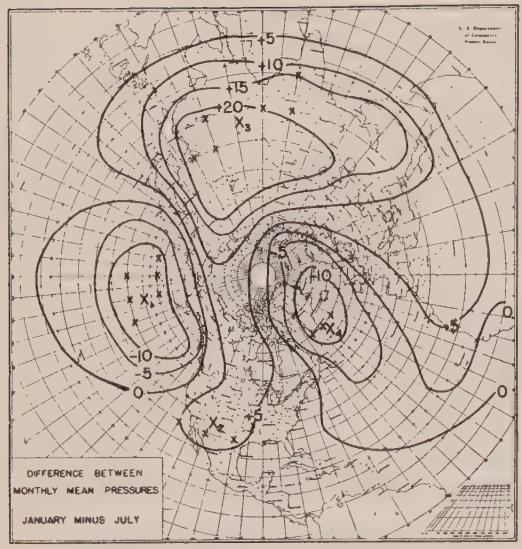


Fig. 2. Isopleths of difference in normal sea level pressure in mbs. between January and July for the Northern Hemisphere.

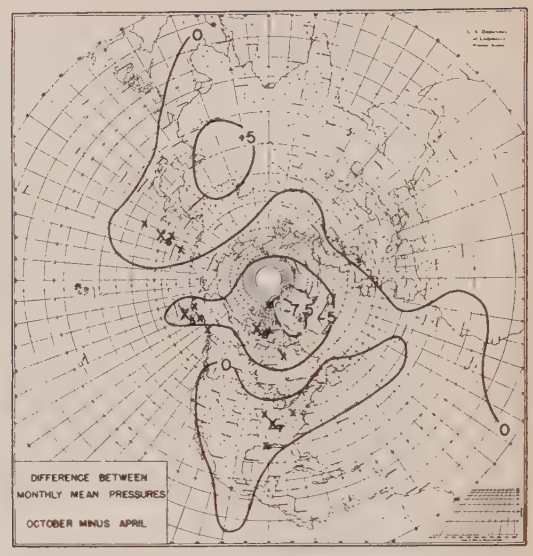


Figure 3. The difference in normal sea level pressure between October and April for the Northern Hemisphere.

statistical analysis of such an index might disclose interesting facts. Likewise, since Figure 3 indicates differences between October and April on the average, it would appear possible to determine another index V which would indicate the extent to which a particular monthly map resembled a spring map compared with a fall map.

On the basis of Figure 2, four regions were selected where the average pressure was computed for each month of the series. The points selected are shown as crosses on the figure and the areas are labelled X_1, X_2, X_3 and X_4.

When the average pressure indices X_1 and X_2 were plotted against each other the points for January fell in the upper left side of the graph and the points for July fell in the lower right hand section. The points for the other months fell in between but with considerable overlap between adjacent months. The means for each month are shown in Figure 4, but the points for the individual months are not reproduced here. Using a graphical regression technique described elsewhere (Brier, 1946) a series of arbitrary lines Y_1 indicated as curves in Figure 4 were drawn so that all the points had values of Y_1 ranging between 0 and 100. The spacing and curvature of these lines was determined by the data for the individual months which are not plotted here. The values of Y_1 were then read off for each individual month of the 40 year series and plotted against the index X_3 on a chart similar to that shown in Figure 5. Lines Z_1 ranging between 0 and 100 were similarly constructed on this figure as shown. For each month, the value of Z_1 was read off the chart and a final index W was computed by the formula

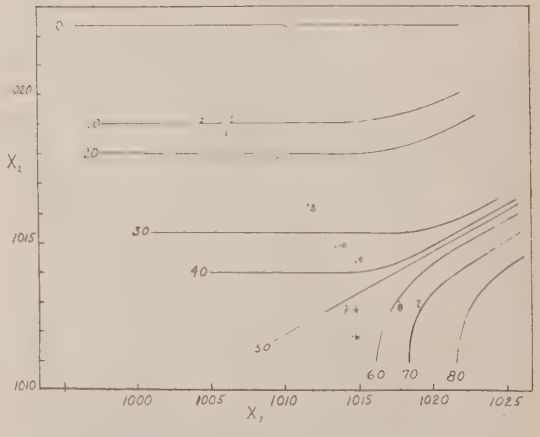


Fig. 4. Scatter diagram showing the separation of the twelve monthly normal maps with respect to variables X_1 and X_2. Months are numbered in order from 1 for January to 12 for December. The curves are indices of resemblance to the July normal chart.

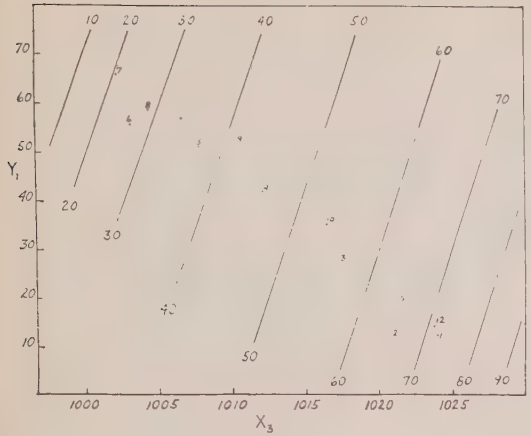


Fig. 5. Scatter diagram showing the separation of the twelve monthly normal maps with respect to variables X_3 and Y_4. The slanting lines are indices of resemblance to the January normal chart. Months are numbered in order from 1 for January to 12 for December.

$$W = \frac{Z_1 - \bar{Z}_1}{\sigma_{z_1}}$$

where $$\bar{Z}_1 = \sum_i^{486} Z_{1i}/486$$

and $$\sigma_{z_1} = \sqrt{\frac{\Sigma (Z_{1i} - \bar{Z})^2}{486}}$$

Thus, the index W has been scaled to have a mean of zero and standard deviation of unity. The average values of W for each of the twelve months are shown in Table 2. An attempt was made to use variable X_4, but it was found to contribute no additional discriminating information.

For the spring-fall index, V, the regions chosen are shown in Figure 3. After a little experimentation it was found that the best discrimination was obtained when Y_3 was plotted against Y_4 where

$$Y_3 = X_5 - X_6,$$
$$Y_4 = X_7 - X_8.$$

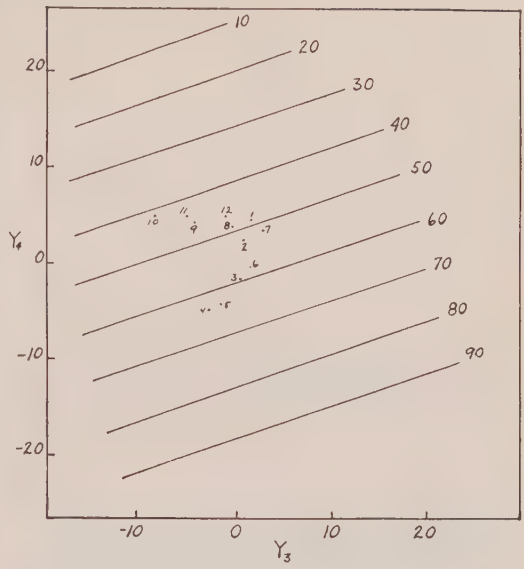


Fig. 6. Scatter diagram showing the separation of the twelve monthly normal maps with respect to variables Y_3 and Y_4. The slanting lines are indices of resemblance to the April normal chart. Months are numbered in order from 1 for January to 12 for December.

After Y_3 was plotted against Y_4, the lines Z_2 were drawn as shown in Figure 6. When the individual monthly values of Z_2 were determined from their position on Figure 6 and tabulated they were found to have a slight correlation with index W. In order to have an index independent of W the linear regression of Z_2 on W was determined and a final index V was computed by the formula

$$V = \frac{Z_2 - \bar{Z}_2 + bW}{\sigma_r}$$

where

$$\bar{Z}_2 = \sum_i^{486} Z_{2i}/486$$

and b is the regression of Z_2 on W and σ_r is the standard deviation of the residuals from the regression line. Thus, index V has zero mean and unit standard deviation and is

Table 1. Average values of index W for each month.

Month	Jan.	Feb.	Mar.	Apr.	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.
Index W....	1.31	1.01	.41	— .16	— .65	— 1.25	— 1.61	— 1.13	— .41	.28	.88	1.26

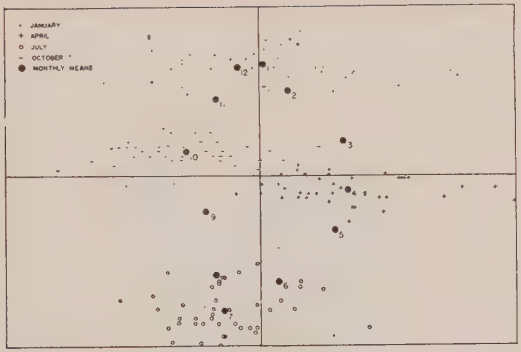


Fig. 7. Scatter diagram showing how the W and V indices separate the January maps from the July maps and the April maps from the October maps. Months are numbered in order from 1 for January to 12 for December.

uncorrelated with W since it is a residual from the regression line of Z_2 on W.

The effectiveness of these two indices in separating the winter from the summer maps and the spring maps from the fall maps is indicated by Figure 7. On this figure the mean values of W and V are shown for each of the twelve months of the year but the individual monthly values are plotted only for January, April, July and October. If the remaining individual months were shown they would indicate considerable scatter and overlap between adjacent calendar months.

4. Statistical Analysis of the Indices

A number of correlations were computed between W and V and other factors. The

Table 2. Linear Trends in W and V.

Month	Correlation Coefficient		Regression Coefficient	
	W	V	W	V
January	— 0.45	— 0.35	— 0.0068	— 0.029
February ...	— 0.59	— 0.44	— 0.0090	— 0.032
March.....	— 0.60	— 0.23	— 0.0058	— 0.016
April.....	— 0.64	— 0.07	— 0.0079	— 0.004
May	— 0.55	— 0.21	— 0.0071	— 0.015
June.....	— 0.47	— 0.32	— 0.0072	— 0.019
July	— 0.49	— 0.18	— 0.0119	— 0.009
August.....	— 0.25	— 0.30	— 0.0046	— 0.017
September..	— 0.63	— 0.05	— 0.0065	— 0.002
October	— 0.44	— 0.08	— 0.0051	— 0.005
November ..	— 0.56	— 0.39	— 0.0072	— 0.033
December ..	— 0.45	— 0.33	— 0.0084	— 0.023
Annual	— 0.75	— 0.36	— 0.0732	— 0.070

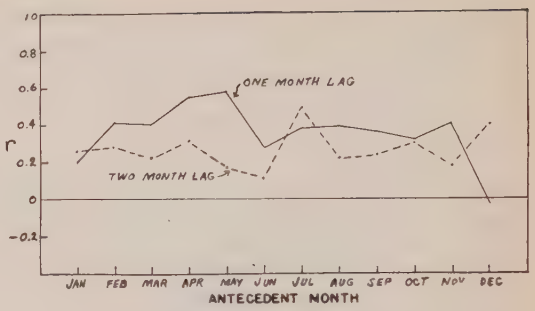


Fig. 8. Auto-correlations of index W for one and two month lags.

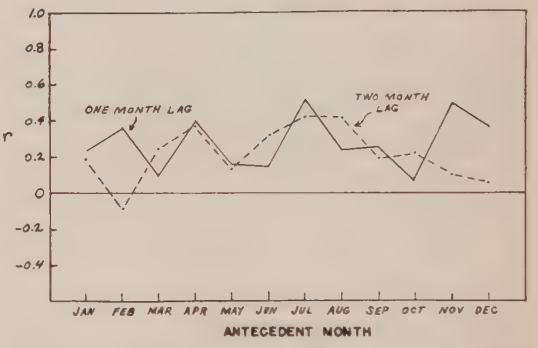


Fig. 9. Auto-correlations of index V for one and two month lags.

persistence in these indices was investigated by computing the one and two month lag autocorrelation for each month of the year. The results of these computations are shown graphically in Figures 8 and 9. The one month lag correlations were highest for W with values greater than 0.50 for April with May and for May with June. The average of the one month lag correlations was 0.35 for W and 0.27 for V. Ordinarily, these would not be considered as very useful for forecasting purposes.

The trends in the indices during the period were investigated by computing the regression on time for each index and for each month. The results, shown in Table 2, indicate that during the 40 years there were pronounced trends in both of these indices and that the sign of the trend was the same in all seasons. Since positive values of W means more winter-like indices, the negative correlations mean that on the average the monthly mean patterns were more like normal winter pat-

terns in the earlier part of the period and more like the normal summer pressure patterns toward the end of the period. The most pronounced trend was during the month of April and Figure 10 shows a plot of the individual values of W for the 41 months of April and the regression line on time. This tendency for the maps to show more summer-like patterns during the 40 year period appears to be consistent with the findings of AHLMANN (1948), KINCER (1946) and other investigators reporting a general warming up of the hemisphere during this period. The negative correlations for index V are not so easy to interpret but indicate a tendency for the monthly mean pressure patterns to become more like normal fall patterns during the 40 years.

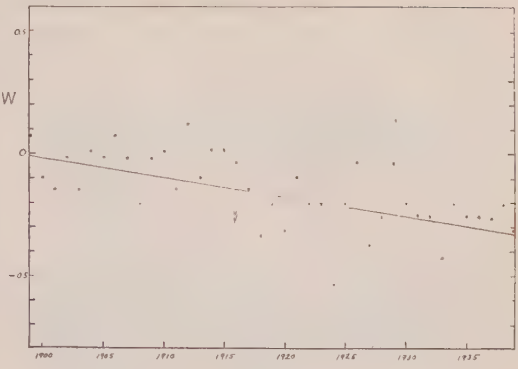


Fig. 10. Time series plot of index W for the month of April.

Index W was determined on the basis of areas showing the greatest change in pressure from January to July. Since the basic cause of this seasonal change is undoubtedly radiational in character, due to the changing position of the sun relative to the earth, it would seem reasonable to hypothesize that real changes in solar activity might affect the same regions. Therefore, the index W was correlated with sunspots (MUNRO 1948) the easiest available measure of solar activity; the results for each of the twelve months are shown by the solid line in Figure 11. Correlations between V and sunspots were also computed although there was no a priori reason to expect any relationship. The solid line in Figure 12 shows a graph of these correlations.

For index W, all twelve of the correlations turned out negative, ranging in value from -0.07 to -0.44. Using the standard test of significance, two of these would be considered significant at the 1 % level and another one at the 5 % level. The algebraic average of the twelve monthly correlations is -0.26 and on the basis of the t test this is a significant departure from zero. For index V the situation is quite different. None of the twelve monthly correlations is significant at the 5 % level, and the average of the group does not depart significantly from zero.

Since the analysis indicated a considerable trend in W and V during the 40-½ year period, the question might be raised as to whether the observed negative correlations between W and sunspot numbers might only be the result of a trend in the sunspot numbers opposite the trend in W. For example, the trend in W during the period might have no physical reality, resulting from some bias in the data introduced by the analysis of the original synoptic maps which were based on less complete data in the earlier years. Since sunspots showed a slight upward trend during the period, a "spurious" correlation between them and W would be found even though no physical relationship existed between the two. For this reason the correlations between sunspots and the indices were recomputed after removing the linear trends in W, V and sunspots. The results are shown graphically by the dashed lines in Figures 11 and 12.

The removal of the trends in the series reduced the correlations between W and sunspots only one of them (October) remaining significant at the 5 % level. However, 10 of the 12 correlations are of the same sign as before and the average for the group of -0.10 departs significantly from zero. The fact that the correlations of W and sunspots are negative after the trends are removed and that the trend in sunspots is opposite to the trend in W strongly suggest that the correlations found between sunspots and W presented by the solid line in Figure 11 are not spurious. Furthermore, it is the author's opinion that the trends in index W during the period were real and cannot be explained by asserting that insufficient data in the earlier years biased the results in a considerable way. The correlations of sunspots with V show essentially the same thing as before the removal of trends, the group average of 0.05 not being

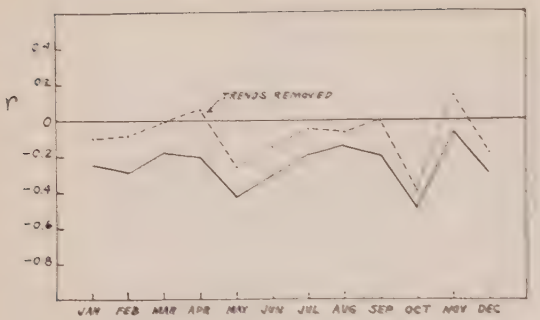


Fig. 11. Correlations of index W with sunspots, before and after removal of the trend during the period.

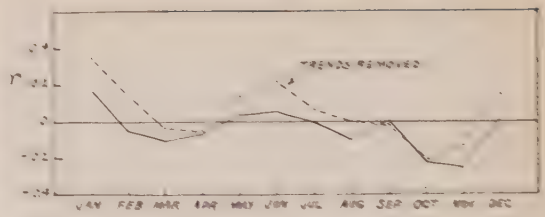


Fig. 12. Correlations of index V with sunspots, before and after removal of the trend during the period.

significant although the January correlation of 0.34 is significant at the 5 % level.

The validity of some of the results reported here might be open to some question due to the fact that some of the points used in determining the indices W and V are located in regions of less reliable data and weather analysis. For example, variable X_2 includes a point located in the mountains of Mexico and another point located in the Pacific off Lower California. This question was investigated by substituting for variable X_2 the mean monthly pressure data for Corpus Christi, Phoenix, and San Diego. The correlations of these stations with variable X_2 averaged 0.92. Likewise, the correlations between variable X_3 and a new variable X_3' which eliminated questionable points located in China and Tibet ranged between 0.87 and 0.94. These changes were too small to have an appreciable effect on the W and V indices and their relationships to the sunspot series, and indicate that the significant correlations reported here cannot be simply explained by the assertion they were produced by faulty data or map analysis.

5. Conclusions

A method of map classification has been developed which can be used to type monthly mean sea-level pressure maps according to the degree to which they resemble a normal winter or summer circulation pattern on one hand and a normal spring or fall map on the other. Index W based on the average pressure of points chosen from three regions measures the resemblance to the January normal map and index V is a measure of the resemblance to the normal spring map.

Indices W and V show considerable persistence from one to two months with autocorrelations ranging from —0.03 to 0.58 for one month lags and from —0.09 to 0.51 for two month lags. Some of these correlations appear to be large and consistent enough to suggest a possible usefulness in long range forecasting.

Both indices show rather consistent negative trends during the period, indicating that the monthly mean patterns were more like summer and fall normal pressure patterns toward the end of the period and more like winter and spring during the earlier part of the period. These results seem to be consistent with the findings of others investigating the climatic flucatuations of the past half-century.

For each month the correlation of index W with sunspots was negative, indicating that periods of high sunspots tended to be associated with summer type patterns, at least so far as these are determined by the distribution of pressure in regions chosen for this study. Since most of the correlations remained negative when the trend was removed from both the sunspots and W series, there appears to be the definite suggestion that part of the variation in the monthly mean sea level pressure patterns during the period 1899 to 1939 was produced by variable solar activity. The correlations are not large enough to indicate that a very substantial portion of variation was solar in origin and they are not high enough to be of much forecasting value, even if they are real. However, the results appear to give some slight supporting evidence to the conclusion reached by Willett (1940) that irregular solar activity is responsible for some of the changes in the circulation of the atmosphere.

REFERENCES

AHLMAN, H. W., 1948: The Present Climatic Fluctuation, *Geographical Journal*, vol. **62**. 4—6, pp. 165—195.

BRIER, G. W., 1946: A study of quantitative precipitation forecasting in the TVA Basin, *U. S. Weather Bureau research paper No. 26*, 40 pp.

BRIER, G. W., 1947: Forty year trends in Northern Hemisphere surface pressure, *Bull. Amer. Met. Soc.*, **28**, pp. 237—247.

BRIER, G. W., 1948: Changes in the Northern Hemisphere sea-level circulation in the 40 year period 1899—1939. *Transactions, AGU*, **29**, 6, pp. 789—795.

KINCER, J. B., 1946: Our changing Climate. *Transactions AGU*, **27**, 3, pp. 342—347.

MUNRO, ELLA H., 1948: Tables on Sunspot-Frequency for 1749—1948, *Journal of Geophysical Research*, **53**, 3.

U. S. Weather Bureau and Army Air Forces, 1943—44: *Historical Weather Maps, Northern Hemisphere, Sea Level.*

U. S. Weather Bureau and Army Air Forces 1946, a: *Normal Weather Maps, Northern Hemisphere, Sea Level Pressure.*

U. S. Weather Bureau, 1946, b: *Mean Monthly Pressure Maps, Northern Hemisphere Sea Level*, (Unpublished).

WILLETT, H. C., 1949, Long-period fluctuations of the general circulation of the atmosphere. *Journal of Meteorology*, **6**, 1, pp. 34—50.

CORRIGENDUM

November issue of Tellus 1951

B. BOLIN, J. G. CHARNEY: Numerical Tendency Computations from the Barotropic Vorticity Equation.

Due to an oversight the article did not contain an acknowledgement that the work involved was performed under contract N6-ori, Task Order 1, between the Office of Naval Research and the Institute for Advanced Study. The authors would like to make this acknowledgement now.

B. BOLIN J. CHARNEY